



Stability of finite difference schemes for hyperbolic initial boundary value problems

Jean-François Coulombel

► To cite this version:

Jean-François Coulombel. Stability of finite difference schemes for hyperbolic initial boundary value problems. SIAM Journal on Numerical Analysis, 2009, 47 (4), pp.2844-2871. hal-00339527v2

HAL Id: hal-00339527

<https://hal.science/hal-00339527v2>

Submitted on 12 Jan 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Stability of finite difference schemes for hyperbolic initial boundary value problems

Jean-François COULOMBEL

CNRS & Université Lille 1, Laboratoire Paul Painlevé (UMR CNRS 8524)

and Project Team SIMPAF of INRIA Lille Nord Europe

Cité scientifique, Bâtiment M2, 59655 VILLENEUVE D'ASCQ Cedex, France

Email: jfcoulom@math.univ-lille1.fr

January 12, 2009

Abstract

We study the stability of finite difference schemes for hyperbolic initial boundary value problems in one space dimension. Assuming ℓ^2 -stability for the discretization of the hyperbolic operator as well as a geometric regularity condition, we show that an appropriate determinant condition, that is the analogue of the uniform Kreiss-Lopatinskii condition for the continuous problem, yields strong stability for the discretized initial boundary value problem. The analysis relies on a suitable *discrete block structure condition* and the construction of suitable symmetrizers. Our work extends the results of [9] to a wider class of finite difference schemes.

AMS subject classification: 65N12, 65N06, 35L50.

Keywords: Hyperbolic systems, boundary conditions, finite difference schemes, stability, symmetrizers.

1 Introduction

The goal of this article is to make a precise study of the stability of finite difference schemes for hyperbolic initial boundary value problems in one space dimension. There has been a wide series of works on this subject, see e.g. [10, 18, 9, 5, 6] and the references therein. In the fundamental contribution [9], it was shown that strong stability of finite difference approximations is equivalent to an appropriate “determinant condition”, under the assumption that the discretization of the hyperbolic operator is either dissipative or unitary, see Assumption 5.4 in [9]. The determinant condition is the analogue of the uniform Kreiss-Lopatinskii condition for hyperbolic boundary value problems, see [11]. The results of [9] were extended to the multidimensional case in [16], assuming that the discretization of the hyperbolic operator is dissipative in the tangential directions. We also refer to [3] for a study of multidimensional finite volume approximations in the case of symmetric systems.

In one space dimension, the dissipativity assumption is not very restrictive. As a matter of fact, many finite difference approximations of the equation:

$$\partial_t u + A \partial_x u = f, \quad (t, x) \in \mathbb{R}^+ \times \mathbb{R},$$

are dissipative under the assumption that 0 is not an eigenvalue of A and that the CFL condition is not satisfied in an optimal way. For initial boundary value problems on the half-line $\{x > 0\}$,

this assumption on the matrix A means precisely that the boundary $\{x = 0\}$ is noncharacteristic. In two space dimensions, the dissipativity assumption is much more restrictive (it is also much more difficult to check). In some cases, finite difference approximations of the equation:

$$\partial_t u + A_1 \partial_{x_1} u + A_2 \partial_{x_2} u = f, \quad (t, x_1, x_2) \in \mathbb{R}^+ \times \mathbb{R} \times \mathbb{R},$$

can be dissipative only if neither A_1 nor A_2 have 0 as an eigenvalue, which is far more restrictive than assuming that the boundary of the half-space $\{x_2 > 0\}$ is non-characteristic. It is also rather restrictive from the point of view of applications (one can think of the linearized gas dynamics equations for instance).

The purpose of this article is therefore to extend the theory of [9] to the widest possible class of finite difference schemes. The reasons are twofold: first to cover as many applications as possible by diminishing the assumptions on the finite difference scheme, second to highlight the structural assumptions that are needed in the procedure to derive maximal energy estimates for the discretization of the initial boundary value problem. Our goal is also to avoid as much as possible using the specificities of hyperbolic systems in one space dimension so our results will be useful for a future extension to the multidimensional case. We shall give two examples of schemes that are covered by our results and that do not enter the framework of [9]. Let us now briefly recall the procedure to derive energy estimates for hyperbolic boundary value problems.

In the analysis of hyperbolic initial boundary value problems (both for continuous and discretized problems), energy estimates are based on two main steps. The first step consists in writing the problem as an “evolution equation” in the normal direction to the boundary, and in reducing the symbol of this evolution equation under a convenient form (that is usually called the block structure). Usually this evolution equation in the normal variable is not hyperbolic so the symbol cannot be reduced to diagonal form. The second step of the analysis is to construct symmetrizers for this reduced form of the equations. The “compatibility” between the operator and the boundary conditions is encoded in a determinant condition, usually known as the uniform Kreiss-Lopatinskii condition. For the continuous problem, the block structure condition was proved by Kreiss [11] for strictly hyperbolic operators and non-characteristic boundary. Majda and Osher [12] observed that this condition is satisfied by many physical examples that are hyperbolic with constant multiplicity, both in the characteristic and non-characteristic case. The derivation of the determinant condition and the construction of symmetrizers were performed in [11]. Kreiss’s result was extended by Métivier [13] to hyperbolic operators with constant multiplicity (both for noncharacteristic and characteristic boundaries). Eventually, it was shown by Métivier and Zumbrun [15] that the block structure condition is equivalent to a geometric regularity property for the eigenvalues of the symbol associated with the Cauchy problem. The result in [15] characterizes completely the structural conditions on the hyperbolic operator that make the symmetrizers construction in [11] work.

Our goal is to extend the results of [15] to the “discrete” case. This requires first of all to define a suitable *discrete block structure condition*, then to show that geometric regularity of the eigenvalues of the symbol of the discretized hyperbolic operator is equivalent to the discrete block structure condition. The next step of the analysis is to construct symmetrizers and to prove maximal energy estimates. In the discrete case, the eigenvalues of the symbol of the discretized hyperbolic operator can have a more complex behavior than the eigenvalues of the symbol of the continuous hyperbolic operator. Consequently, we have less information about the diagonal blocks involved in the discrete block structure. We are then led to make some restrictions on the discretized hyperbolic operator that give some additional information on the blocks and allow us to construct a symmetrizer. We shall consider a wider class of schemes than those considered in [9]. As a matter of fact, the restrictions in [9] were such that the authors could use the

symmetrizers constructed in [11]. Weakening the assumptions of [9] requires the introduction of a new type of symmetrizers. The construction of these new symmetrizers is performed here in details. We also improve the results of [9] by showing some refined results under weaker assumptions. Our new construction of symmetrizers is flexible enough to be generalized to even more general situations than what we consider here (we give a possible generalization in an appendix).

Notations In all this paper, we use the notations:

$$\begin{aligned}\mathcal{U} &:= \{\zeta \in \mathbb{C}, |\zeta| > 1\}, \quad \overline{\mathcal{U}} := \{\zeta \in \mathbb{C}, |\zeta| \geq 1\}, \\ \mathbb{D} &:= \{\zeta \in \mathbb{C}, |\zeta| < 1\}, \quad \mathbb{S}^1 := \{\zeta \in \mathbb{C}, |\zeta| = 1\}.\end{aligned}$$

We let $\mathcal{M}_{p,N}(\mathbb{K})$ denote the set of $p \times N$ matrices with entries in $\mathbb{K} = \mathbb{R}$ or \mathbb{C} , and we use the notation $\mathcal{M}_N(\mathbb{K})$ when $p = N$. If $M \in \mathcal{M}_N(\mathbb{C})$, $\text{sp}(M)$ denotes the spectrum of M , while M^* denotes the conjugate transpose of M . The matrix $(M + M^*)/2$ is called the real part of M and is denoted $\text{Re } M$. We let I denote the identity matrix, without mentioning the dimension. If $H_1, H_2 \in \mathcal{M}_N(\mathbb{C})$ are two hermitian matrices, we write $H_1 \geq H_2$ if for all $x \in \mathbb{C}^N$ we have $x^*(H_1 - H_2)x \geq 0$. The norm of a vector $x \in \mathbb{C}^N$ is $|x| := (x^*x)^{1/2}$. Eventually, we let ℓ^2 denote the set of square integrable sequences, without mentioning the indices of the sequences (sequences may be valued in \mathbb{C}^k for some integer k).

2 Main results

We consider a hyperbolic initial boundary value problem in one space dimension:

$$\begin{cases} \partial_t u + A \partial_x u = F(t, x), & (t, x) \in \mathbb{R}^+ \times \mathbb{R}^+, \\ B u(t, 0) = g(t), & t \in \mathbb{R}^+, \\ u(0, x) = f(x), & x \in \mathbb{R}^+, \end{cases} \quad (1)$$

where $A \in \mathcal{M}_N(\mathbb{R})$ is diagonalizable with real eigenvalues, and $B \in \mathcal{M}_{N_+, N}(\mathbb{R})$ with N_+ the number of positive eigenvalues of A (counted with their multiplicity). We assume that the boundary is noncharacteristic, that is $0 \notin \text{sp}(A)$. Problem (1) is well-posed in L^2 if and only if:

$$\mathbb{R}^N = \text{Ker } B \oplus E_+(A),$$

where $E_+(A)$ is the unstable eigenspace of A (associated with positive eigenvalues of A).

We now introduce the finite difference approximation of (1). Let $\Delta x, \Delta t > 0$ denote a space and a time step where $\lambda = \Delta t / \Delta x$ is a fixed positive constant, and let p, q, r, s be some integers. The solution to (1) is approximated by a sequence (U_j^n) defined for $n \in \mathbb{N}$, and $j \in -r+1+\mathbb{N}$. For $j = -r+1, \dots, 0$, U_j^n approximates the trace $u(n\Delta t, 0)$ on the boundary $\{x = 0\}$, and possibly the trace of normal derivatives. The boundary meshes $[j\Delta x, (j+1)\Delta x]$, $j = -r+1, \dots, 0$, shrink to $\{0\}$ as Δx tends to 0, so the “formal” continuous limit problem as Δx tends to 0 is

set on the half-line \mathbb{R}^+ . We consider finite difference approximations of (1) that read¹:

$$\begin{cases} U_j^{n+1} = \sum_{\sigma=0}^s Q_\sigma U_j^{n-\sigma} + \Delta t F_j^n, & j \geq 1, \quad n \geq s, \\ U_j^{n+1} = \sum_{\sigma=-1}^s B_{j,\sigma} U_1^{n-\sigma} + g_j^n, & j = -r+1, \dots, 0, \quad n \geq s, \\ U_j^n = f_j^n, & j \geq -r+1, \quad n = 0, \dots, s, \end{cases} \quad (2)$$

where the operators Q_σ and $B_{j,\sigma}$ are given by:

$$Q_\sigma := \sum_{\ell=-r}^p A_{\ell,\sigma} T^\ell, \quad B_{j,\sigma} := \sum_{\ell=0}^q B_{\ell,j,\sigma} T^\ell, \quad T^\ell U_k^m := U_{k+\ell}^m. \quad (3)$$

In (3), all matrices $A_{\ell,\sigma}, B_{\ell,j,\sigma}$ belong to $\mathcal{M}_N(\mathbb{R})$. We recall the following definition from [9]:

Definition 1 (Strong stability [9]). *The finite difference approximation (2) is said to be strongly stable if there exists a constant C such that for all $\gamma > 0$ and all $\Delta t \in]0, 1]$, the solution (U_j^n) of (2) with $f_j^n = 0$ satisfies the estimate:*

$$\begin{aligned} & \frac{\gamma}{\gamma \Delta t + 1} \sum_{n \geq s} \sum_{j \geq -r+1} \Delta t \Delta x e^{-2\gamma n \Delta t} |U_j^n|^2 + \sum_{n \geq s} \sum_{j=-r+1}^0 \Delta t e^{-2\gamma n \Delta t} |U_j^n|^2 \\ & \leq C \left\{ \frac{\gamma \Delta t + 1}{\gamma} \sum_{n \geq s} \sum_{j \geq 1} \Delta t \Delta x e^{-2\gamma n \Delta t} |F_j^n|^2 + \sum_{n \geq s} \sum_{j=-r+1}^0 \Delta t e^{-2\gamma n \Delta t} |g_j^n|^2 \right\}. \end{aligned}$$

The estimate in definition 1 is the discrete counterpart of the maximal energy estimate for the “continuous” problem (1):

$$\begin{aligned} & \gamma \iint_{\mathbb{R}^+ \times \mathbb{R}^+} e^{-2\gamma t} |u(t, x)|^2 dt dx + \int_{\mathbb{R}^+} e^{-2\gamma t} |u(t, 0)|^2 dt \\ & \leq C \left\{ \frac{1}{\gamma} \iint_{\mathbb{R}^+ \times \mathbb{R}^+} e^{-2\gamma t} |F(t, x)|^2 dt dx + \int_{\mathbb{R}^+} e^{-2\gamma t} |g(t)|^2 dt \right\}. \end{aligned}$$

For later use, we introduce the symbol associated with the discretization of the hyperbolic operator:

$$\forall \kappa \in \mathbb{C} \setminus \{0\}, \quad \mathcal{A}(\kappa) := \begin{pmatrix} \widehat{Q}_0(\kappa) & \dots & \dots & \widehat{Q}_s(\kappa) \\ I & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & I & 0 \end{pmatrix} \in \mathcal{M}_{N(s+1)}(\mathbb{C}), \quad \widehat{Q}_\sigma(\kappa) := \sum_{\ell=-r}^p \kappa^\ell A_{\ell,\sigma}. \quad (4)$$

The uniform power boundedness of $\mathcal{A}(\kappa)$ for $\kappa \in \mathbb{S}^1$ is a necessary and sufficient condition to have ℓ^2 -stability for the discretized Cauchy problem, see for instance [4, chapter III.1] or [8, chapter 5]. This stability condition for the discretized Cauchy problem will play an important role in the stability analysis for the discretized initial boundary value problem (2), just like

¹We do not focus here on the construction of such approximations and refer to [8] for some examples that enter our framework.

hyperbolicity plays an important role for the stability analysis of hyperbolic initial boundary value problems, see e.g. [11, 13, 15].

Let us now introduce the resolvent equation:

$$\begin{cases} w_j - \sum_{\sigma=0}^s z^{-\sigma-1} Q_\sigma w_j = F_j, & j \geq 1, \\ w_j - \sum_{\sigma=-1}^s z^{-\sigma-1} B_{j,\sigma} w_1 = g_j, & j = -r+1, \dots, 0, \end{cases} \quad (5)$$

where $z \in \mathcal{U}$, $(F_j) \in \ell^2$, and $g_{-r+1}, \dots, g_0 \in \mathbb{C}^N$, that is obtained from (2) by applying a Laplace transform in time, see [9]. Then strong stability of (2) can be first characterized by an estimate on the resolvent equation:

Proposition 1 ([9]). *The approximation (2) is strongly stable if and only if there exists a constant $C > 0$ such that for all $z \in \mathcal{U}$, for all $(F_j) \in \ell^2$, and for all $g_{-r+1}, \dots, g_0 \in \mathbb{C}^N$, the resolvent equation (5) has a unique solution $(w_j) \in \ell^2$ and this solution satisfies:*

$$\frac{|z|-1}{|z|} \sum_{j \geq -r+1} |w_j|^2 + \sum_{j=-r+1}^0 |w_j|^2 \leq C \left\{ \frac{|z|}{|z|-1} \sum_{j \geq 1} |F_j|^2 + \sum_{j=-r+1}^0 |g_j|^2 \right\}. \quad (6)$$

The goal of this article is to give necessary and/or sufficient conditions on the symbol (4) and on the boundary conditions in (2) so that the energy estimate (6) holds true, which implies strong stability in the sense of definition 1. As detailed in the introduction, it is convenient to rewrite the resolvent equation (5) as an “evolution equation” for the sequence (w_j) . For $\ell = -r, \dots, p$, we define the matrices:

$$\forall z \in \mathbb{C} \setminus \{0\}, \quad \mathbb{A}_\ell(z) := \delta_{\ell 0} I - \sum_{\sigma=0}^s z^{-\sigma-1} A_{\ell,\sigma}, \quad (7)$$

where $\delta_{\ell_1 \ell_2}$ is the Kronecker symbol. Then as in [9], we make the following assumption:

Assumption 1. *The matrices $\mathbb{A}_{-r}(z)$ and $\mathbb{A}_p(z)$ are invertible for all $z \in \overline{\mathcal{U}}$, or equivalently for all z in some open neighborhood \mathcal{V} of \mathcal{U} .*

As usual, it is convenient to rewrite the “multi-step” induction (5) as a “one-step” induction for an augmented vector. Assumption 1 is crucial to achieve this reduction.

We first consider the case $q < p$. In that case, all the w_j ’s involved in the “boundary conditions” for the resolvent equation (5) are coordinates of the augmented vector² $W_1 := (w_p, \dots, w_{1-r})$. Using assumption 1, we can define a matrix $\mathbb{M}(z)$ that is holomorphic on some open neighborhood \mathcal{V} of \mathcal{U} :

$$\forall z \in \mathcal{V}, \quad \mathbb{M}(z) := \begin{pmatrix} -\mathbb{A}_p(z)^{-1} \mathbb{A}_{p-1}(z) & \dots & \dots & -\mathbb{A}_p(z)^{-1} \mathbb{A}_{-r}(z) \\ I & 0 & \dots & 0 \\ 0 & \ddots & \ddots & \vdots \\ 0 & 0 & I & 0 \end{pmatrix} \in \mathcal{M}_{N(p+r)}(\mathbb{C}). \quad (8)$$

²Vectors are written indifferently in rows or columns to simplify the redaction.

Using the definition (3) for the operators $Q_\sigma, B_{j,\sigma}$ and the matrix \mathbb{M} , we can rewrite the resolvent equation (5) as an induction relation for the augmented vector $W_j := (w_{j+p-1}, \dots, w_{j-r})$. This induction relation reads:

$$\begin{cases} W_{j+1} = \mathbb{M}(z) W_j + \widetilde{F}_j, & j \geq 1, \\ B(z) W_1 = g, \end{cases} \quad (9)$$

where the new source terms $(\widetilde{F}_j), g$ in (9) are given by:

$$\widetilde{F}_j := (\mathbb{A}_p(z)^{-1} F_j, 0, \dots, 0), \quad g := (g_0, \dots, g_{1-r}).$$

It is easy to check that the matrix $B(z) \in \mathcal{M}_{Nr, N(p+r)}$ depends holomorphically on $z \in \mathbb{C} \setminus \{0\}$ and has maximal rank Nr for all z . The exact expression of the matrix $B(z)$ can be easily obtained from (5) and (3) but is not very relevant here. The new resolvent equation (9) is of course equivalent to (5).

Let us now treat the case $q \geq p$. In that case, we can still write the resolvent equation under the form (9) up to defining $W_j := (w_{j+q}, \dots, w_{j-r}), j \geq 1$, and:

$$\mathbb{M}(z) := \begin{pmatrix} -\mathbb{A}_p(z)^{-1} \mathbb{A}_{p-1}(z) & \dots & -\mathbb{A}_p(z)^{-1} \mathbb{A}_{-r}(z) & 0 & \dots & 0 \\ I & 0 & \dots & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & I & 0 \\ 0 & 0 & \dots & 0 & I & 0 \end{pmatrix} \in \mathcal{M}_{N(q+r+1)}(\mathbb{C}).$$

The definition of $B(z) \in \mathcal{M}_{Nr, N(q+r+1)}$ varies from the previous case but this matrix keeps a maximal rank Nr for all z and is still holomorphic on $\mathbb{C} \setminus \{0\}$. This equivalent form of the resolvent equation varies from what was done in [9]. In our approach, we can ensure that the matrix $B(z)$ has maximal rank for all $z \in \overline{\mathcal{U}}$. This is important in view of the so-called uniform Kreiss-Lopatinskii condition assumed below in Theorem 2. This maximal rank property of B is not so clear if one uses the method in [9, page 672] when $q \geq p$. This is why we propose this alternative approach to rewrite the resolvent equation.

For simplicity, we shall deal from now on with the case $q < p$ but our proofs can be extended in a straightforward way to the case $q \geq p$. Some intermediate results vary slightly but the method and arguments are the same in both cases.

Our goal is to prove an energy estimate for (9), which will yield an energy estimate for (5) and prove strong stability for the finite difference scheme (2) thanks to Proposition 1. We introduce the following terminology:

Definition 2 (Discrete block structure condition). *Let M be a holomorphic function on some open neighborhood of $\overline{\mathcal{U}}$ with values in $\mathcal{M}_m(\mathbb{C})$ for some integer m . Then M is said to satisfy the discrete block structure condition if the two following conditions are satisfied:*

1. *for all $z \in \mathcal{U}$, $\text{sp}(M(z)) \cap \mathbb{S}^1 = \emptyset$,*
2. *for all $\underline{z} \in \overline{\mathcal{U}}$, there exists an open neighborhood \mathcal{O} of \underline{z} in \mathbb{C} , and there exists an invertible matrix $T(z)$ that is holomorphic with respect to $z \in \mathcal{O}$ such that:*

$$\forall z \in \mathcal{O}, \quad T(z)^{-1} M(z) T(z) = \text{diag} (M_1(z), \dots, M_L(z)),$$

where the number L of diagonal blocks and the size ν_ℓ of each block M_ℓ do not depend on $z \in \mathcal{O}$, and where each block satisfies one of the following properties:

- *there exists $\delta > 0$ such that for all $z \in \mathcal{O}$, $M_\ell(z)^* M_\ell(z) \geq (1 + \delta) I$,*

- there exists $\delta > 0$ such that for all $z \in \mathcal{O}$, $M_\ell(z)^* M_\ell(z) \leq (1 - \delta) I$,
- $\nu_\ell = 1$, \underline{z} and $M_\ell(\underline{z})$ belong to \mathbb{S}^1 , and $\underline{z} M'_\ell(\underline{z}) \overline{M_\ell(\underline{z})} \in \mathbb{R} \setminus \{0\}$,
- $\nu_\ell > 1$, $\underline{z} \in \mathbb{S}^1$ and $M_\ell(\underline{z})$ has the form:

$$M_\ell(\underline{z}) = \underline{\kappa}_\ell \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & \dots & 0 & 1 \end{pmatrix}, \quad \underline{\kappa}_\ell \in \mathbb{S}^1.$$

Moreover the lower left coefficient m_ℓ of $M'_\ell(\underline{z})$ is such that for all $\mu \in \mathbb{C}$ with $\operatorname{Re} \sigma > 0$, and for all complex number ζ such that $\zeta^{\nu_\ell} = \overline{\underline{\kappa}_\ell} m_\ell \underline{z} \mu$, then $\operatorname{Re} \zeta \neq 0$.

We refer to the blocks M_ℓ as being of the first, second, third or fourth type.

The discrete block structure condition is more precise than the normal form of [9, Theorem 9.1]. Definition 2 clarifies the structure of the blocks associated with eigenvalues in \mathbb{S}^1 . Such blocks are either scalar, which was not clear in [9], or have a Jordan structure. This clarification will simplify the construction of symmetrizers.

Our first main result gives necessary and sufficient conditions on the symbol \mathcal{A} so that the matrix \mathbb{M} defined by (8) satisfies the discrete block structure condition:

Theorem 1. *Let assumption 1 be satisfied. Then \mathbb{M} defined by (8) satisfies the discrete block structure condition if and only if the symbol \mathcal{A} defined by (4) satisfies the two following conditions:*

- there exists a constant $C > 0$ such that for all $\kappa \in \mathbb{S}^1$ and all $n \in \mathbb{N}$, $\|\mathcal{A}(\kappa)^n\| \leq C$, (uniform power boundedness)
- if $\underline{\kappa} \in \mathbb{S}^1$ and $\underline{z} \in \mathbb{S}^1 \cap \operatorname{sp}(\mathcal{A}(\underline{\kappa}))$ has algebraic multiplicity $\underline{\alpha}$, then there exist some functions $\lambda_1(\kappa), \dots, \lambda_{\underline{\alpha}}(\kappa)$ that are holomorphic in a neighborhood \mathcal{W} of $\underline{\kappa}$ in \mathbb{C} , such that:

$$\begin{aligned} \lambda_1(\underline{\kappa}) &= \dots = \lambda_{\underline{\alpha}}(\underline{\kappa}) = \underline{z}, \\ \det(z I - \mathcal{A}(\kappa)) &= \vartheta(\kappa, z) \prod_{j=1}^{\underline{\alpha}} (z - \lambda_j(\kappa)), \end{aligned} \tag{10}$$

with ϑ a holomorphic function of (κ, z) in some neighborhood of $(\underline{\kappa}, \underline{z})$ in \mathbb{C}^2 such that $\vartheta(\underline{\kappa}, \underline{z}) \neq 0$, and there exist some vectors $E_1(\kappa), \dots, E_{\underline{\alpha}}(\kappa) \in \mathbb{C}^{N(s+1)}$ that are holomorphic with respect to $\kappa \in \mathcal{W}$, that are linearly independent for all $\kappa \in \mathcal{W}$, and that satisfy:

$$\forall \kappa \in \mathcal{W}, \quad \forall j = 1, \dots, \underline{\alpha}, \quad \mathcal{A}(\kappa) E_j(\kappa) = \lambda_j(\kappa) E_j(\kappa).$$

Theorem 1 shows that \mathbb{M} satisfies the discrete block structure condition if and only if the discretization of the hyperbolic operator is ℓ^2 -stable and moreover the eigenvalues of the symbol \mathcal{A} that belong to the unit circle are geometrically regular. No geometric regularity is required for eigenvalues in \mathbb{D} . Theorem 1 is the analogue for finite difference schemes of Theorem C.3 in [15]. The assumptions of Theorem 1 allow more general situations than the cases covered by [9]. In particular, we show that assumptions 5.2 and 5.3 in [9] are not necessary to reduce \mathbb{M} to the discrete block structure. Let us make the following:

Remark 1. We recall that the geometric regularity of the eigenvalues in \mathbb{S}^1 is not a consequence of the uniform power boundedness of $\{\mathcal{A}(\kappa), \kappa \in \mathbb{S}^1\}$. For instance the following matrix:

$$\begin{pmatrix} 1 + \Upsilon(\kappa) & \Upsilon(\kappa) \\ 0 & 1 + \Upsilon(\kappa) \end{pmatrix}, \quad \Upsilon(\kappa) := \frac{(\kappa - \kappa^{-1})^2}{4},$$

is holomorphic with respect to $\kappa \in \mathbb{C} \setminus \{0\}$, and is uniformly power bounded for $\kappa \in \mathbb{S}^1$. However 1 is not a geometrically regular eigenvalue for $\kappa = 1$.

The reduction of \mathbb{M} to the discrete block structure is a crucial step towards proving energy estimates as we shall see below. It is indeed easier to construct symmetrizers for each block rather than directly constructing a symmetrizer for the whole matrix \mathbb{M} .

In what follows, we are going to give sufficient conditions on the symbol \mathcal{A} and on the boundary conditions in (2) that yield strong stability. Unfortunately, our structural conditions on \mathcal{A} are slightly more restrictive than the conditions in Theorem 1 that ensure the discrete block condition for \mathbb{M} . However we shall allow more general situations than what was considered in [9]. We shall also give some examples of new situations where our analysis applies.

Let us introduce the following terminology that is inspired from [14]:

Definition 3 (*K-symmetrizer*). Let assumption 1 be satisfied, and let \mathbb{M} be defined by (8). Then \mathbb{M} is said to admit a *K-symmetrizer* if for all $\underline{z} \in \overline{\mathcal{U}}$, there exists a decomposition:

$$\mathbb{C}^{N(p+r)} = \underline{\mathbb{E}}^s \oplus \underline{\mathbb{E}}^u, \quad \dim \underline{\mathbb{E}}^s = N r, \quad \dim \underline{\mathbb{E}}^u = N p,$$

with associated projectors (π^s, π^u) , such that for all $K \geq 1$, there exists a neighborhood \mathcal{O} of \underline{z} in \mathbb{C} , there exists a \mathcal{C}^∞ function S on \mathcal{O} with values in $\mathcal{M}_{N(p+r)}(\mathbb{C})$, and there exists a constant $c > 0$ such that the following properties hold for all $z \in \mathcal{O} \cap \mathcal{U}$:

- $S(z)$ is hermitian,
- $\mathbb{M}(z)^* S(z) \mathbb{M}(z) - S(z) \geq c(|z| - 1)/|z| I$,
- for all $W \in \mathbb{C}^{N(p+r)}$, $W^* S(z) W \geq K^2 |\pi^u W|^2 - |\pi^s W|^2$.

Before going on, we make the following:

Remark 2. In definition 3, the neighborhood \mathcal{O} , the mapping S and the constant c heavily depend on K . What is important is that the decomposition of $\mathbb{C}^{N(p+r)}$ only depends on the point \underline{z} that is considered.

If $\underline{z} \in \mathcal{U}$, the first estimate in definition 3 is equivalent to proving that $\mathbb{M}(\underline{z})^* S(\underline{z}) \mathbb{M}(\underline{z}) - S(\underline{z})$ is positive definite.

The last estimate in definition 3 can be deduced from the estimate:

$$\forall W \in \mathbb{C}^{N(p+r)}, \quad W^* S(\underline{z}) W \geq \left(K^2 + \frac{1}{2}\right) |\pi^u W|^2 - \frac{1}{2} |\pi^s W|^2.$$

This reduces the third point to a verification only at the point \underline{z} . We shall repeatedly use this property in our construction of *K-symmetrizers*.

Our second main result gives sufficient conditions for strong stability of (2) and is the following:

Theorem 2. *Let assumption 1 be satisfied. Assume that the symbol $\mathcal{A}(\kappa)$ defined by (4) is uniformly power bounded for $\kappa \in \mathbb{S}^1$, and that \mathbb{M} defined by (8) admits a K -symmetrizer. For $z \in \mathcal{U}$, we let $\mathbb{E}^s(z)$ denote the generalized eigenspace associated with eigenvalues of $\mathbb{M}(z)$ in \mathbb{D} . Then $\mathbb{E}^s(z)$ has constant dimension Nr for all $z \in \mathcal{U}$ and \mathbb{E}^s defines a holomorphic vector bundle over \mathcal{U} that can be extended in a unique way as a continuous vector bundle over $\overline{\mathcal{U}}$. We let $\mathbb{E}^s(z)$ denote this continuous extension for $z \in \mathbb{S}^1 (= \partial\mathcal{U})$.*

In addition to all assumptions above, assume that for all $z \in \overline{\mathcal{U}}$ we have $\mathbb{E}^s(z) \cap \text{Ker } B(z) = \{0\}$. (In what follows this condition is referred to as the uniform Kreiss-Lopatinskii condition.) Then the scheme (2) is strongly stable.

Theorem 2 reduces the verification of the strong stability of (2) to *i)* constructing a suitable symmetrizer for \mathbb{M} in the neighborhood of any point $z \in \overline{\mathcal{U}}$, and *ii)* verifying the so-called uniform Kreiss-Lopatinskii condition. This condition may be formulated as a *determinant condition* provided that we choose a basis of $\mathbb{E}^s(z)$ and a basis of $\text{Ker } B(z)$. Point *i)* is usually performed by first reducing \mathbb{M} to the discrete block structure, and this is where Theorem 1 is useful. The verification of the uniform Kreiss-Lopatinskii condition may be quite involved since it requires to compute the continuous extension of \mathbb{E}^s to $\overline{\mathcal{U}}$. For low order schemes where r is small, the computations may be performed with no major difficulty, see e.g. [9, section 6]; however the verification becomes really complicated for high order schemes with large r . Nevertheless, one may try a numerical verification of this condition, as indicated in [8, chapter 13]. The continuous extension of $\mathbb{E}^s(z)$ to $z \in \overline{\mathcal{U}}$ is not explicitly proved in [9] though it is crucially needed in the formulation of the so-called determinant condition.

Our last main result gives sufficient conditions for the existence of a K -symmetrizer. Combined with the result of Theorem 2 above, Theorem 3 below thus gives sufficient conditions for strong stability of (2).

Theorem 3. *Let assumption 1 be satisfied, and assume that the symbol $\mathcal{A}(\kappa)$ is uniformly power bounded for $\kappa \in \mathbb{S}^1$. Assume moreover that all the eigenvalues of $\mathcal{A}(\kappa)$, $\kappa \in \mathbb{S}^1$, that belong to \mathbb{S}^1 are geometrically regular and that at least one of the three following properties is satisfied by each eigenvalue $\lambda_j(\kappa)$ in the decomposition (10):*

- i) $\lambda_j'(\underline{\kappa}) \neq 0$,*
- ii) $\lambda_j(\kappa) \in \mathbb{S}^1$ for all $\kappa \in \mathbb{S}^1 \cap \mathcal{W}$,*
- iii) $\text{Re}(\underline{\kappa}^2 \overline{\lambda_j(\underline{\kappa})} \lambda_j''(\underline{\kappa})) > 0$.*

Then \mathbb{M} defined by (8) admits a K -symmetrizer. In particular, if the uniform Kreiss-Lopatinskii condition is satisfied, the scheme (2) is strongly stable.

The rest of this paper is devoted to the proofs of Theorems 1, 2 and 3. In section 6 we shall detail some examples of schemes for which our analysis applies. This section includes new examples that were not covered by the analysis of [9]. Appendices A and B are devoted to the proof of intermediate results that are used in the proofs of the main Theorems. Eventually, appendix C is devoted to an extension of Theorem 3 where we consider even more general situations. The purpose of appendix C is to convince the reader that our new construction of symmetrizers can be adapted to any situation where we make a dissipation type assumption for the eigenvalues λ_j . A systematic treatment in the general case is postponed to a future work.

Possible generalizations of our results are detailed in the following:

Remark 3. A possible generalization of Theorem 2 is to show that the vector bundle \mathbb{E}^s can be continuously extended to $\overline{\mathcal{U}}$ assuming only that the discrete block structure condition is satisfied by \mathbb{M} . This would allow to define the uniform Kreiss-Lopatinskii condition as in Theorem 2 for a wider class of numerical schemes. However, checking the uniform Kreiss-Lopatinskii condition is useless if one is not also able to construct symmetrizers. In our work, extending continuously the bundle \mathbb{E}^s to $\overline{\mathcal{U}}$ appears as a consequence of the symmetrizers construction, even though it may be already hidden in the discrete block structure condition.

Our construction of symmetrizers is smooth with respect to the frequency z . We thus have all the ingredients for an extension of our results to variable coefficients difference operators. The only new ingredient that is needed is a suitable quantification for pseudo-difference operators, see e.g. [16, section 4].

Eventually, our results generalize the theory in [9] so we can extend to our more general framework some previous results that were based on the results of [9], see e.g. [5, 6, 7] for simplified stability criteria or convergence estimates. Since we have not used the specificities of the one-dimensional framework, our analysis can be a good starting point for a generalization of Michelson's results [16] in several space dimensions. This is left to a future work.

3 Proof of Theorem 1

Let us recall that we assume $q < p$ for simplicity. We first assume that the symbol $\mathcal{A}(\kappa)$ is uniformly power bounded for $\kappa \in \mathbb{S}^1$ and that its eigenvalues that belong to the unit circle \mathbb{S}^1 are geometrically regular. We are going to show that \mathbb{M} satisfies the discrete block structure condition. Recall that \mathbb{M} is holomorphic on an open neighborhood \mathcal{V} of $\overline{\mathcal{U}}$, see (8). We first recall a classical Lemma:

Lemma 1 ([9]). *Let assumption 1 be satisfied and assume that $\mathcal{A}(\kappa)$ is uniformly power bounded for $\kappa \in \mathbb{S}^1$. Then for all $z \in \mathcal{V}$, the eigenvalues of $\mathbb{M}(z)$ are those $\kappa \in \mathbb{C} \setminus \{0\}$ such that:*

$$\det(\mathcal{A}(\kappa) - zI) = 0.$$

In particular for all $z \in \mathcal{U}$, $\mathbb{M}(z)$ has no eigenvalue on the unit circle \mathbb{S}^1 and the number of eigenvalues in \mathbb{D} equals Nr (eigenvalues are counted with their algebraic multiplicity).

Lemma 1 shows that the first condition in definition 2 is satisfied. Furthermore, this property immediately implies that the discrete block structure condition is satisfied in the neighborhood of any $\underline{z} \in \mathcal{U}$. More precisely, in a small neighborhood \mathcal{O} of $\underline{z} \in \mathcal{U}$, the generalized eigenspace associated with eigenvalues of $\mathbb{M}(z)$ in \mathbb{D} and the generalized eigenspace associated with eigenvalues of $\mathbb{M}(z)$ in \mathcal{U} both depend holomorphically on $z \in \mathcal{O}$. We can then reduce $\mathbb{M}(z)$ to a block diagonal form:

$$T(z)^{-1} \mathbb{M}(z) T(z) = \text{diag}(\mathbb{M}_b(z), \mathbb{M}_\sharp(z)), \quad \mathbb{M}_b(z) \in \mathcal{M}_{Nr}(\mathbb{C}), \quad \mathbb{M}_\sharp(z) \in \mathcal{M}_{Np}(\mathbb{C}),$$

where the eigenvalues of $\mathbb{M}_b(z)$ belong to \mathbb{D} and the eigenvalues of $\mathbb{M}_\sharp(z)$ belong to \mathcal{U} . The change of basis $T(z)$ depends holomorphically on $z \in \mathcal{O}$. Up to a constant change of basis, we can achieve the inequalities:

$$\mathbb{M}_b(\underline{z})^* \mathbb{M}_b(\underline{z}) \leq (1 - 2\delta)I, \quad \mathbb{M}_\sharp(\underline{z})^* \mathbb{M}_\sharp(\underline{z}) \geq (1 + 2\delta)I,$$

with δ some positive constant. Thanks to a continuity argument, we can conclude that the discrete block structure condition is satisfied in a neighborhood of \underline{z} . The reduction only involves blocks of the first and second type.

We now turn to the case $\underline{z} \in \mathbb{S}^1$. If $\mathbb{M}(\underline{z})$ has no eigenvalue in \mathbb{S}^1 then we are reduced to the preceding case. We thus assume that $\mathbb{M}(\underline{z})$ has some eigenvalues in \mathbb{S}^1 . More precisely, let $\underline{\kappa}_1, \dots, \underline{\kappa}_k$ denote the elements of $\text{sp}(\mathbb{M}(\underline{z})) \cap \mathbb{S}^1$. We also let $\alpha_1, \dots, \alpha_k$ denote the algebraic multiplicity of these eigenvalues. The generalized eigenspace $\text{Ker}(\mathbb{M}(\underline{z}) - \underline{\kappa}_j I)^{\alpha_j}$ associated with $\underline{\kappa}_j$ is denoted \mathbb{K}_j . For z sufficiently close to \underline{z} , we also let $\mathbb{K}_j(z)$ denote the generalized eigenspace of $\mathbb{M}(z)$ associated with its α_j eigenvalues that are close to $\underline{\kappa}_j$. The space $\mathbb{K}_j(z)$ depends holomorphically on z . Then for z in a small neighborhood \mathcal{O} of \underline{z} , we can perform a block diagonalization of $\mathbb{M}(z)$ with a holomorphic change of basis:

$$T(z)^{-1} \mathbb{M}(z) T(z) = \text{diag} (\mathbb{M}_b(z), \mathbb{M}_\sharp(z), \mathbb{M}_1(z), \dots, \mathbb{M}_k(z)),$$

where the eigenvalues of $\mathbb{M}_b(z)$ belong to \mathbb{D} , the eigenvalues of $\mathbb{M}_\sharp(z)$ belong to \mathcal{U} , and for all $j = 1, \dots, k$ the α_j eigenvalues of $\mathbb{M}_j(z)$ belong to a sufficiently small neighborhood of $\underline{\kappa}_j$. As in the preceding case, we can easily achieve the inequalities:

$$\forall z \in \mathcal{O}, \quad \mathbb{M}_b(z)^* \mathbb{M}_b(z) \leq (1 - \delta) I, \quad \mathbb{M}_\sharp(z)^* \mathbb{M}_\sharp(z) \geq (1 + \delta) I,$$

so from now on we focus on the blocks $\mathbb{M}_j(z)$. For the sake of clarity, we shall only deal with the block $\mathbb{M}_1(z)$. This is only to avoid overloaded notations with many indices. Of course, the analysis below is valid for any of the blocks $\mathbb{M}_j(z)$. We are going to show that in a convenient holomorphic basis of $\mathbb{K}_1(z)$, the block $\mathbb{M}_1(z)$ reduces to a block diagonal form with blocks of the third and fourth types. The proof follows the analysis of [13, 15] and splits in several steps.

• Following [13], we first study the characteristic polynomial of $\mathbb{M}_1(z)$. For z close to \underline{z} , the α_1 eigenvalues of $\mathbb{M}_1(z)$ are close to $\underline{\kappa}_1$. A standard computation shows that for all $z \in \mathcal{V}$ and all $\kappa \in \mathbb{C} \setminus \{0\}$, we have³:

$$\det(\mathbb{M}(z) - \kappa I) = \vartheta(\kappa, z) \det(z I - \mathcal{A}(\kappa)), \quad (11)$$

where ϑ is holomorphic with respect to (κ, z) and does not vanish on $\mathbb{C} \setminus \{0\} \times \mathcal{V}$. We thus obtain:

$$\det(\mathbb{M}_1(z) - \kappa I) = \vartheta(\kappa, z) \det(z I - \mathcal{A}(\kappa)), \quad (12)$$

where ϑ is holomorphic with respect to (κ, z) and does not vanish on a small neighborhood of $(\underline{\kappa}_1, \underline{z})$. The generic notation ϑ is used for a nonvanishing holomorphic function that may vary from one line to the other. We know that $\underline{z} \in \mathbb{S}^1$ is an eigenvalue of $\mathcal{A}(\underline{\kappa}_1)$ so we can use the geometric regularity assumption. For (κ, z) in a sufficiently small neighborhood of $(\underline{\kappa}_1, \underline{z})$, (12) reads:

$$\det(\mathbb{M}_1(z) - \kappa I) = \vartheta(\kappa, z) \prod_{j=1}^{\underline{\alpha}} (z - \lambda_j(\kappa)), \quad (13)$$

where $\underline{\alpha}$ is a fixed integer, and the λ_j 's are holomorphic functions on a neighborhood \mathcal{W} of $\underline{\kappa}_1$ and satisfy $\lambda_j(\underline{\kappa}_1) = \underline{z}$. Thanks to the uniform power boundedness of the matrices $\mathcal{A}(\kappa)$ for $\kappa \in \mathbb{S}^1$, we know that $|\lambda_j(\kappa)| \leq 1$ for $\kappa \in \mathbb{S}^1 \cap \mathcal{W}$. Using a Taylor expansion of $\lambda_j(\underline{\kappa}_1 e^{i\xi})$ for $\xi \in \mathbb{R}$ close to 0, we obtain that there exists a real number α_j such that:

$$\underline{\kappa}_1 \lambda_j'(\underline{\kappa}_1) = \alpha_j \underline{z}, \quad \alpha_j \in \mathbb{R}. \quad (14)$$

Thanks to (13), we can see that $\underline{\kappa}_1$ is a root with finite multiplicity ν_j of the holomorphic function $\underline{z} - \lambda_j(\cdot)$:

$$\forall \nu = 1, \dots, \nu_j - 1, \quad \lambda_j^{(\nu)}(\underline{\kappa}_1) = 0, \quad \lambda_j^{(\nu_j)}(\underline{\kappa}_1) \neq 0. \quad (15)$$

³This relation is a little stronger than the result recalled in Lemma 1, but it can be obtained by applying standard computation rules for determinants.

We can therefore apply the Weierstrass preparation Theorem to the holomorphic function $z - \lambda_j(\kappa)$: for all $j = 1, \dots, \underline{\alpha}$, there exists $P_j(\kappa, z)$ that is a unitary polynomial function in κ with degree ν_j , such that for (κ, z) close to $(\underline{\kappa}_1, \underline{z})$:

$$z - \lambda_j(\kappa) = \vartheta(\kappa, z) P_j(\kappa, z), \quad P_j(\kappa, \underline{z}) = (\kappa - \underline{\kappa}_1)^{\nu_j}, \quad \vartheta(\underline{\kappa}_1, \underline{z}) \neq 0. \quad (16)$$

Using (16), (13) reduces to:

$$\det(\mathbb{M}_1(z) - \kappa I) = \vartheta(\kappa, z) \prod_{j=1}^{\underline{\alpha}} P_j(\kappa, z). \quad (17)$$

For z close to \underline{z} , the polynomial $P_j(\cdot, z)$ has ν_j roots close to $\underline{\kappa}_1$. Consequently, the size of the block $\mathbb{M}_1(z)$ equals $\nu_1 + \dots + \nu_{\underline{\alpha}}$. We also know that the size of this block equals α_1 , the algebraic multiplicity of the eigenvalue $\underline{\kappa}_1$. Up to reordering the terms, there exists an integer $\underline{\beta}$ such that:

$$\nu_1 = \dots = \nu_{\underline{\beta}} = 1, \quad \nu_{\underline{\beta}+1}, \dots, \nu_{\underline{\alpha}} \geq 2.$$

For $j = 1, \dots, \underline{\beta}$, we know that $\lambda'_j(\underline{\kappa}_1) \neq 0$ or equivalently $\alpha_j \neq 0$ in (14). Therefore λ_j is a biholomorphic homeomorphism from a neighborhood \mathscr{W} of $\underline{\kappa}_1$ to a neighborhood \mathscr{O} of \underline{z} . We let μ_j denote its (holomorphic) inverse. With such notations, we obtain $P_j(\kappa, z) = \kappa - \mu_j(z)$ for all $j = 1, \dots, \underline{\beta}$.

Using the relation (16), we also obtain $\partial_z P_j(\underline{\kappa}_1, \underline{z}) \neq 0$. Then Puiseux's expansions theory shows that for z close to \underline{z} and $z \neq \underline{z}$, the ν_j roots of $P_j(\cdot, z)$ are simple, see for instance [1].

• For each eigenvalue $\lambda_j(\kappa)$, $j = 1, \dots, \underline{\alpha}$ and κ close to $\underline{\kappa}_1$, we know that $\mathscr{A}(\kappa)$ has a holomorphic eigenvector $E_j(\kappa) \in \mathbb{C}^{N(s+1)}$. Using the definition (4) of \mathscr{A} , we find that $E_j(\kappa)$ can be written as:

$$\forall j = 1, \dots, \underline{\alpha}, \quad E_j(\kappa) = \begin{pmatrix} \lambda_j(\kappa)^s e_j(\kappa) \\ \vdots \\ \lambda_j(\kappa) e_j(\kappa) \\ e_j(\kappa) \end{pmatrix}, \quad e_j(\kappa) \in \mathbb{C}^N.$$

The vectors $e_1(\underline{\kappa}_1), \dots, e_{\underline{\alpha}}(\underline{\kappa}_1)$ are linearly independent in \mathbb{C}^N because $E_1(\underline{\kappa}_1), \dots, E_{\underline{\alpha}}(\underline{\kappa}_1)$ are linearly independent in $\mathbb{C}^{N(s+1)}$. Therefore when κ is close to $\underline{\kappa}_1$, the vectors $e_1(\kappa), \dots, e_{\underline{\alpha}}(\kappa)$ remain linearly independent. We define the following vectors:

$$\forall j = 1, \dots, \underline{\alpha}, \quad \mathscr{E}_j(\kappa) := \begin{pmatrix} \kappa^{p+r-1} e_j(\kappa) \\ \vdots \\ \kappa e_j(\kappa) \\ e_j(\kappa) \end{pmatrix} \in \mathbb{C}^{N(p+r)}.$$

These vectors depend holomorphically on κ , and they are linearly independent for κ close to $\underline{\kappa}_1$. Using the relations (7) and (8), some straightforward computations show that $\mathscr{E}_j(\kappa)$ is an eigenvector of the matrix $\mathbb{M}(\lambda_j(\kappa))$ associated with the eigenvalue κ :

$$\forall j = 1, \dots, \underline{\alpha}, \quad (\mathbb{M}(\lambda_j(\kappa)) - \kappa I) \mathscr{E}_j(\kappa) = 0. \quad (18)$$

In particular, for $j = 1, \dots, \underline{\beta}$ and for z in a neighborhood \mathscr{O} of \underline{z} , we have:

$$\forall j = 1, \dots, \underline{\beta}, \quad \forall z \in \mathscr{O}, \quad (\mathbb{M}(z) - \mu_j(z) I) \mathscr{E}_j(\mu_j(z)) = 0. \quad (19)$$

Recall that μ_j is the holomorphic inverse of λ_j for $j = 1, \dots, \underline{\beta}$, that is when $\lambda'_j(\underline{\kappa}_1) \neq 0$. For all $j = 1, \dots, \underline{\beta}$, we have thus constructed a holomorphic eigenvalue $\mu_j(z)$ and a holomorphic eigenvector $\mathcal{E}_j(\mu_j(z))$ of $\mathbb{M}(z)$. Moreover, we have $\mu'_j(\underline{z}) = 1/\lambda'_j(\underline{\kappa}_1)$ so we get:

$$\forall j = 1, \dots, \underline{\beta}, \quad \mu_j(\underline{z}) = \underline{\kappa}_1 \in \mathbb{S}^1, \quad \underline{z} \mu'_j(\underline{z}) \overline{\mu_j(\underline{z})} = \frac{1}{\alpha_j} \in \mathbb{R} \setminus \{0\}.$$

• We now turn to the most difficult case $j = \underline{\beta} + 1, \dots, \underline{\alpha}$. We start from (18), differentiate this relation $\nu_j - 1$ times with respect to κ , and evaluate the result at $\kappa = \underline{\kappa}_1$:

$$\begin{aligned} (\mathbb{M}(\underline{z}) - \underline{\kappa}_1 I) \mathcal{E}_j(\underline{\kappa}_1) &= 0, \\ -\mathcal{E}_j(\underline{\kappa}_1) + (\mathbb{M}(\underline{z}) - \underline{\kappa}_1 I) \mathcal{E}'_j(\underline{\kappa}_1) &= 0, \\ &\vdots \\ -(\nu_j - 1) \mathcal{E}_j^{(\nu_j-2)}(\underline{\kappa}_1) + (\mathbb{M}(\underline{z}) - \underline{\kappa}_1 I) \mathcal{E}_j^{(\nu_j-1)}(\underline{\kappa}_1) &= 0. \end{aligned}$$

Then for all $j = \underline{\beta} + 1, \dots, \underline{\alpha}$, we define the following vectors:

$$(\underline{\mathcal{E}}_{j,1}, \dots, \underline{\mathcal{E}}_{j,\nu_j}) := \left(\mathcal{E}_j(\underline{\kappa}_1), \frac{\underline{\kappa}_1}{1!} \mathcal{E}'_j(\underline{\kappa}_1), \dots, \frac{\underline{\kappa}_1^{\nu_j-1}}{(\nu_j-1)!} \mathcal{E}_j^{(\nu_j-1)}(\underline{\kappa}_1) \right), \quad (20)$$

that satisfy the relations:

$$(\mathbb{M}(\underline{z}) - \underline{\kappa}_1 I) \underline{\mathcal{E}}_{j,1} = 0, \quad \forall \mu = 2, \dots, \nu_j, \quad (\mathbb{M}(\underline{z}) - \underline{\kappa}_1 I) \underline{\mathcal{E}}_{j,\mu} = \underline{\kappa}_1 \underline{\mathcal{E}}_{j,\mu-1}. \quad (21)$$

Using the relations (19) and (21), we can show as in [13, 15] that the vectors:

$$\mathcal{E}_1(\underline{\kappa}_1), \dots, \mathcal{E}_{\underline{\beta}}(\underline{\kappa}_1), \quad \underline{\mathcal{E}}_{\underline{\beta}+1,1}, \dots, \underline{\mathcal{E}}_{\underline{\beta}+1,\nu_{\underline{\beta}+1}}, \quad \dots, \quad \underline{\mathcal{E}}_{\underline{\alpha},1}, \dots, \underline{\mathcal{E}}_{\underline{\alpha},\nu_{\underline{\alpha}}},$$

are linearly independent. Moreover, these α_1 vectors span the generalized eigenspace \mathbb{K}_1 of $\mathbb{M}(\underline{z})$ associated with the eigenvalue $\underline{\kappa}_1$. So far we have thus obtained a basis of \mathbb{K}_1 in which the block $\mathbb{M}_1(\underline{z})$ reads:

$$\mathbb{M}_1(\underline{z}) = \text{diag} \left(\underline{\kappa}_1, \dots, \underline{\kappa}_1, \underline{M}_{\underline{\beta}+1}, \dots, \underline{M}_{\underline{\alpha}} \right), \quad \underline{M}_j := \underline{\kappa}_1 \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & \ddots & \ddots & 1 \\ 0 & \dots & 0 & 1 \end{pmatrix} \in \mathcal{M}_{\nu_j}(\mathbb{C}).$$

In the final step of the analysis, we are going to extend the definition of the vectors $\underline{\mathcal{E}}_{j,\mu}$ to a neighborhood of \underline{z} . The proof follows the arguments of [13].

• Let us recall that for all $j = 1, \dots, \underline{\alpha}$, the polynomial P_j is defined by (16). We can choose $r > 0$ such that for z in a neighborhood \mathcal{O} of \underline{z} , the ν_j roots of $P_j(\cdot, z)$ belong to the disc of center $\underline{\kappa}_1$ and radius $r/2$. Then for all $z \in \mathcal{O}$, for all $j = \underline{\beta} + 1, \dots, \underline{\alpha}$ and for all $\mu = 1, \dots, \nu_j$, we define a vector $\mathcal{E}_{j,\mu}(z)$ by the following formula:

$$\mathcal{E}_{j,\mu}(z) := \frac{\underline{\kappa}_1^{\mu-1} (\nu_j - \mu)!}{2i\pi \nu_j!} \int_{|\kappa - \underline{\kappa}_1| = r} \frac{\partial_\kappa^\mu P_j(\kappa, z)}{P_j(\kappa, z)} \mathcal{E}_j(\kappa) d\kappa.$$

Cauchy's formula shows that for $z = \underline{z}$, $\mathcal{E}_{j,\mu}(\underline{z})$ coincides with the vector $\underline{\mathcal{E}}_{j,\mu}$ defined by (20). Moreover, $\mathcal{E}_{j,\mu}(z)$ depends holomorphically on $z \in \mathcal{O}$. Consequently we can choose the neighborhood \mathcal{O} such that for all $z \in \mathcal{O}$, the vectors:

$$\mathcal{E}_1(\mu_1(z)), \dots, \mathcal{E}_{\underline{\beta}}(\mu_{\underline{\beta}}(z)), \quad \underline{\mathcal{E}}_{\underline{\beta}+1,1}(z), \dots, \underline{\mathcal{E}}_{\underline{\beta}+1,\nu_{\underline{\beta}+1}}(z), \quad \dots, \quad \underline{\mathcal{E}}_{\underline{\alpha},1}(z), \dots, \underline{\mathcal{E}}_{\underline{\alpha},\nu_{\underline{\alpha}}}(z), \quad (22)$$

are linearly independent. We are going to show that these vectors span the invariant subspace $\mathbb{K}_1(z)$, and that in this basis of $\mathbb{K}_1(z)$, the matrix $\mathbb{M}_1(z)$ is in block diagonal form with blocks of the third and fourth type.

We follow [13]: for z close to \underline{z} and $j = \underline{\beta} + 1, \dots, \underline{\alpha}$, we let $\mathbb{F}_j(z)$ denote the vector space spanned by the linearly independent vectors $\mathcal{E}_{j,1}(z), \dots, \mathcal{E}_{j,\nu_j}(z)$. For $j = 1, \dots, \underline{\beta}$, we let $\mathbb{F}_j(z)$ denote the one-dimensional vector space spanned by $\mathcal{E}_j(\mu_j(z))$. Then for all j , the dimension of $\mathbb{F}_j(z)$ is ν_j . Moreover the sum of the $\mathbb{F}_j(z)$ is direct and has dimension α_1 . We already know that for $j = 1, \dots, \underline{\beta}$, $\mathcal{E}_j(\mu_j(z))$ is an eigenvector of $\mathbb{M}(z)$ for the eigenvalue $\mu_j(z)$, see (19). Consequently, $\mathbb{F}_j(z)$ is stable by the matrix $\mathbb{M}(z)$ and $\mathbb{F}_j(z) \subset \mathbb{K}_1(z)$ for $j = 1, \dots, \underline{\beta}$. We are now going to show that the same properties hold true for $j = \underline{\beta} + 1, \dots, \underline{\alpha}$. For $z = \underline{z}$, thanks to (21), we know that $\mathbb{F}_j(\underline{z})$ is stable by $\mathbb{M}(\underline{z})$ and $\mathbb{F}_j(\underline{z}) \subset \mathbb{K}_1$. From now on we thus consider a fixed $z \in \mathcal{O} \setminus \{\underline{z}\}$.

For all $j = \underline{\beta} + 1, \dots, \underline{\alpha}$, we let $\kappa_{j,1}, \dots, \kappa_{j,\nu_j}$ denote the ν_j distinct roots of the polynomial $P_j(\cdot, z)$. These roots belong to the disc of center $\underline{\kappa}_1$ and radius $r/2$. Therefore, using the residue Theorem, we obtain:

$$\mathcal{E}_{j,\mu}(z) = \sum_{m=1}^{\nu_j} \omega_{j,\mu,m} \mathcal{E}_j(\kappa_{j,m}),$$

for some suitable complex numbers $\omega_{j,\mu,m}$. Therefore $\mathbb{F}_j(z)$ is contained in the vector space $\tilde{\mathbb{F}}_j(z)$ spanned by the vectors $\mathcal{E}_j(\kappa_{j,1}), \dots, \mathcal{E}_j(\kappa_{j,\nu_j})$. Because the dimension of $\mathbb{F}_j(z)$ is ν_j , we can conclude that the dimension of $\tilde{\mathbb{F}}_j(z)$ is also ν_j and $\mathbb{F}_j(z) = \tilde{\mathbb{F}}_j(z)$. Let us now show that $\tilde{\mathbb{F}}_j(z)$ is stable by $\mathbb{M}(z)$. We know that $P_j(\kappa_{j,m}, z) = 0$ so $z = \lambda_j(\kappa_{j,m})$. Using (18) we see that $\mathcal{E}_j(\kappa_{j,m})$ is an eigenvector of $\mathbb{M}(z)$ for the eigenvalue $\kappa_{j,m}$ that is close to $\underline{\kappa}_1$. Consequently the vector space $\tilde{\mathbb{F}}_j(z)$ is stable by $\mathbb{M}(z)$ and $\tilde{\mathbb{F}}_j(z) \subset \mathbb{K}_1(z)$. Since $\mathbb{F}_j(z) = \tilde{\mathbb{F}}_j(z)$, we have proved that for all $j = 1, \dots, \underline{\alpha}$, $\mathbb{F}_j(z)$ is stable by $\mathbb{M}(z)$ and $\mathbb{F}_j(z) \subset \mathbb{K}_1(z)$. Using a dimension argument, we have obtained:

$$\mathbb{K}_1(z) = \mathbb{F}_1(z) \oplus \dots \oplus \mathbb{F}_{\underline{\alpha}}(z),$$

and each $\mathbb{F}_j(z)$ is a stable vector space for $\mathbb{M}(z)$. Moreover, the characteristic polynomial of the restriction of $\mathbb{M}(z)$ to $\mathbb{F}_j(z)$ is $P_j(\cdot, z)$. We have thus constructed a holomorphic basis of $\mathbb{K}_1(z)$ in which the matrix $\mathbb{M}_1(z)$ reads:

$$\mathbb{M}_1(z) = \text{diag}(\mu_1(z), \dots, \mu_{\underline{\beta}}(z), M_{\underline{\beta}+1}(z), \dots, M_{\underline{\alpha}}(z)).$$

We also know that the characteristic polynomial of $M_j(z)$ is $P_j(\cdot, z)$ for $j = \underline{\beta} + 1, \dots, \underline{\alpha}$, and $M_j(\underline{z})$ is the Jordan block \underline{M}_j defined above.

• The only remaining task is to obtain the property stated in definition 2 for the lower left corner coefficient m_j of $M'_j(\underline{z})$. We know that $P_j(\kappa, z)$ is the characteristic polynomial of $M_j(z)$. Computing the derivative of $\det(M_j(z) - \kappa_1 I)$ with respect to z and evaluating at $z = \underline{z}$, we obtain $m_j \neq 0$ (because $\partial_z P_j(\underline{\kappa}_1, \underline{z}) \neq 0$). We can then compute the first term in the Puiseux expansion of the eigenvalues of $M_j(z)$ as in [9] (we refer the reader to [1] for a complete justification of the Puiseux expansions for the eigenvalues of $M_j(z)$). Following [9, page 676], the conclusion on m_j follows from the following estimate:

Lemma 2 ([9]). *Let $\mathcal{A}(\kappa)$ be uniformly power bounded for $\kappa \in \mathbb{S}^1$ and let assumption 1 be satisfied. Then there exists a constant $C > 0$ such that for all $z \in \mathcal{U}$ and for all $\kappa \in \mathbb{S}^1$, we have:*

$$\|(\mathbb{M}(z) - \kappa I)^{-1}\| \leq C \frac{|z|}{|z| - 1}.$$

We give an alternative elementary proof of Lemma 2 in appendix A. Our proof relies on very simple algebraic computations and does not use the arguments of [9].

So far we have proved that the uniform power boundedness of $\{\mathcal{A}(\kappa), \kappa \in \mathbb{S}^1\}$ and the geometric regularity of its eigenvalues in \mathbb{S}^1 imply the discrete block structure for \mathbb{M} . From now on, we wish to prove the converse and therefore assume that \mathbb{M} satisfies the discrete block structure assumption. The proof follows [15, appendix C]. First of all, relation (11) shows that for all $\kappa \in \mathbb{S}^1$, the eigenvalues of $\mathcal{A}(\kappa)$ lie in \mathbb{D} . We are first going to show that the eigenvalues of $\mathcal{A}(\kappa)$ that belong to \mathbb{S}^1 are geometrically regular. Then we shall show that $\mathcal{A}(\kappa)$ is uniformly power bounded. We have the following:

Lemma 3. *Let $\mathcal{F} := \{\kappa \in \mathbb{S}^1, \text{sp}(\mathcal{A}(\kappa)) \cap \mathbb{S}^1 \neq \emptyset\}$. Then \mathcal{F} is a closed subset of \mathbb{S}^1 .*

We omit the proof of this Lemma that easily follows from a compactness argument. Let $\underline{\kappa} \in \mathcal{F}$, and let $\underline{z} \in \text{sp}(\mathcal{A}(\underline{\kappa})) \cap \mathbb{S}^1$ have algebraic multiplicity $\underline{\alpha}$. Using relation (11), we know that $\underline{\kappa}$ is an eigenvalue of $\mathbb{M}(\underline{z})$. Because \mathbb{M} satisfies the discrete block structure condition, we know that there exists a neighborhood \mathcal{O} of \underline{z} and a holomorphic change of basis $T(z)$ for $z \in \mathcal{O}$ such that:

$$T(z)^{-1} \mathbb{M}(z) T(z) = \text{diag} (\mathbb{M}_1(z), \dots, \mathbb{M}_{L'}(z), \mathbb{M}_{L'+1}(z), \dots, \mathbb{M}_L(z)),$$

where for all $\ell = 1, \dots, L'$, $\underline{\kappa} \in \text{sp}(\mathbb{M}_\ell(\underline{z}))$ and for all $\ell \geq L' + 1$, $\underline{\kappa} \notin \text{sp}(\mathbb{M}_\ell(\underline{z}))$. Since $\underline{\kappa} \in \mathbb{S}^1$, \mathbb{M}_ℓ is necessarily a block of the third or fourth type⁴ for all $\ell = 1, \dots, L'$. Consequently the spectrum of $\mathbb{M}_\ell(\underline{z})$ is reduced to $\{\underline{\kappa}\}$ for $\ell = 1, \dots, L'$. For all $\ell = 1, \dots, L'$ and $z \in \mathcal{O}$, we define:

$$P_\ell(\kappa, z) = \det(\mathbb{M}_\ell(z) - \kappa I).$$

We clearly have:

$$\det(\mathbb{M}(z) - \kappa I) = \vartheta(\kappa, z) \prod_{\ell=1}^{L'} P_\ell(\kappa, z), \quad \vartheta(\underline{\kappa}, \underline{z}) \neq 0.$$

Using the relation (11), we obtain:

$$\det(z I - \mathcal{A}(\kappa)) = \vartheta(\kappa, z) \prod_{\ell=1}^{L'} P_\ell(\kappa, z), \quad \vartheta(\underline{\kappa}, \underline{z}) \neq 0,$$

and ϑ is holomorphic with respect to (κ, z) in a neighborhood of $(\underline{\kappa}, \underline{z})$. Using the discrete block structure condition, we also have $\partial_z P_\ell(\underline{\kappa}, \underline{z}) \neq 0$ because \mathbb{M}_ℓ is a block of the third or fourth type. Applying the Weierstrass preparation Theorem to P_ℓ , there exists a holomorphic function λ_ℓ defined on a neighborhood \mathcal{W} of $\underline{\kappa}$ such that:

$$P_\ell(\kappa, z) = \vartheta(\kappa, z) (z - \lambda_\ell(\kappa)), \quad \lambda_\ell(\underline{\kappa}) = \underline{z}, \quad \vartheta(\underline{\kappa}, \underline{z}) \neq 0.$$

We have thus obtained the factorization:

$$\det(z I - \mathcal{A}(\kappa)) = \vartheta(\kappa, z) \prod_{\ell=1}^{L'} (z - \lambda_\ell(\kappa)), \quad \vartheta(\underline{\kappa}, \underline{z}) \neq 0.$$

In particular, this shows that L' is the algebraic multiplicity $\underline{\alpha}$ of \underline{z} as an eigenvalue of $\mathcal{A}(\underline{\kappa})$.

⁴Blocks of the first and second type have no eigenvalue in \mathbb{S}^1 .

We know that $\mathbb{M}_\ell(\lambda_\ell(\kappa)) - \kappa I$ has a nontrivial kernel for all κ close to $\underline{\kappa}$ and $\ell = 1, \dots, L'$. Following [15], we can show that this kernel has dimension 1 for all κ , and is spanned by some vector that can be chosen holomorphically with respect to κ . We can therefore construct some vectors $\mathcal{E}_1(\kappa), \dots, \mathcal{E}_{L'}(\kappa)$ that depend holomorphically on κ , that are linearly independent for all $\kappa \in \mathcal{W}$, and that satisfy:

$$\forall \ell = 1, \dots, L', \quad (\mathbb{M}(\lambda_\ell(\kappa)) - \kappa I) \mathcal{E}_\ell(\kappa) = 0.$$

Using some previous computations, the existence of such vectors $\mathcal{E}_\ell(\kappa)$ yields the existence of vectors $E_1(\kappa), \dots, E_{L'}(\kappa) \in \mathbb{C}^{N(s+1)}$ that depend holomorphically on κ , and that satisfy:

$$\forall \ell = 1, \dots, L', \quad (\mathcal{A}(\kappa) - \lambda_\ell(\kappa) I) E_\ell(\kappa) = 0.$$

We have thus proved that \underline{z} is a geometrically regular eigenvalue of $\mathcal{A}(\underline{\kappa})$.

It remains to show that $\mathcal{A}(\kappa)$ is uniformly power bounded. For all $\underline{\kappa} \in \mathcal{F}$, we have shown that there exists a neighborhood \mathcal{W} of $\underline{\kappa}$ in \mathbb{C} , and there exists a holomorphic change of basis $Q(\kappa)$ such that:

$$\forall \kappa \in \mathcal{W}, \quad Q(\kappa)^{-1} \mathcal{A}(\kappa) Q(\kappa) = \text{diag}(\lambda_1(\kappa), \dots, \lambda_L(\kappa), \mathcal{A}_b(\kappa)),$$

where $\lambda_1(\kappa), \dots, \lambda_L(\kappa)$ are complex numbers that satisfy $\lambda_j(\underline{\kappa}) \in \mathbb{S}^1$, and $\mathcal{A}_b(\kappa)$ is a block whose eigenvalues belong to \mathbb{D} for all $\kappa \in \mathcal{W}$. There is no restriction in assuming that \mathcal{W} is open, that $Q(\kappa)$ and $Q(\kappa)^{-1}$ are uniformly bounded on \mathcal{W} , and that the spectrum of $\mathcal{A}_b(\kappa)$ is uniformly bounded away from \mathbb{S}^1 for $\kappa \in \mathcal{W}$.

Since \mathcal{F} is compact, we can use a finite covering of \mathcal{F} by such neighborhoods $\mathcal{W}_1, \dots, \mathcal{W}_k$. For all $\kappa \in \mathbb{S}^1 \cap \mathcal{W}_i$, the eigenvalues $\lambda_1(\kappa), \dots, \lambda_{L_i}(\kappa)$ belong to $\mathbb{D} \cup \mathbb{S}^1$. Moreover, there exists a constant $\delta > 0$ such that for all $\kappa \in \mathcal{W}_i$, the eigenvalues of $\mathcal{A}_b(\kappa)$ have a modulus bounded by $1 - \delta$. Using also a uniform bound for $Q_i(\kappa)$ and $Q_i(\kappa)^{-1}$, we can conclude that there exists a constant $C > 0$ such that:

$$\forall \kappa \in \mathbb{S}^1 \cap \cup_i \mathcal{W}_i, \quad \forall n \in \mathbb{N}, \quad \|\mathcal{A}(\kappa)^n\| \leq C.$$

For κ in the closed subset $(\mathbb{S}^1 \setminus \cup_i \mathcal{W}_i)$ of \mathbb{S}^1 , we know that the spectrum of $\mathcal{A}(\kappa)$ lies inside \mathbb{D} . Consequently, there exists a constant $\delta > 0$ such that for all $\kappa \in (\mathbb{S}^1 \setminus \cup_i \mathcal{W}_i)$, the eigenvalues of $\mathcal{A}(\kappa)$ have a modulus bounded by $1 - \delta$. This shows that there exists a constant $C > 0$ such that:

$$\forall \kappa \in \mathbb{S}^1 \setminus \cup_i \mathcal{W}_i, \quad \forall n \in \mathbb{N}, \quad \|\mathcal{A}(\kappa)^n\| \leq C.$$

The matrix $\mathcal{A}(\kappa)$ is uniformly power bounded for $\kappa \in \mathbb{S}^1$, and the proof of Theorem 1 is complete.

4 Proof of Theorem 2

Let assumption 1 be satisfied, and assume that $\mathcal{A}(\kappa)$ is uniformly power bounded for $\kappa \in \mathbb{S}^1$. Then Lemma 1 shows that the “stable” subspace $\mathbb{E}^s(z)$ of $\mathbb{M}(z)$ has constant dimension Nr for all $z \in \mathcal{U}$. The holomorphic dependance of $\mathbb{M}(z)$ on z implies that $\mathbb{E}^s(z)$ also varies holomorphically with z on \mathcal{U} , see e.g. [1].

Let $\underline{z} \in \mathbb{S}^1$ and let us show that $\mathbb{E}^s(z)$ has a limit as $z \in \mathcal{U}$ tends to \underline{z} . Let $K > 2$, and let us consider the decomposition $\mathbb{C}^{N(p+r)} = \underline{\mathbb{E}}^s \oplus \underline{\mathbb{E}}^u$ as well as the neighborhood \mathcal{O} and the symmetrizer S given in definition 3. We consider $z \in \mathcal{O} \cap \mathcal{U}$ and $W_1 \in \mathbb{E}^s(z)$. We then define the sequence:

$$W_{j+1} = \mathbb{M}(z) W_j, \quad j \geq 1. \tag{23}$$

Because $W_1 \in \mathbb{E}^s(z)$, we have $(W_j) \in \ell^2$. Following some calculations of [9, page 681], we first take the scalar product of (23) with $S(z)W_{j+1}$, take the real part of the equality and sum for $j = 1$ to some integer J :

$$\operatorname{Re} \sum_{j=1}^J W_{j+1}^* S(z) \mathbb{M}(z) W_j - \sum_{j=2}^{J+1} W_j^* S(z) W_j = 0. \quad (24)$$

Then we take the scalar product of (23) with $S(z)\mathbb{M}(z)W_j$, take the real part of the equality and sum for $j = 1$ to J :

$$\sum_{j=1}^J W_j^* \mathbb{M}(z)^* S(z) \mathbb{M}(z) W_j - \operatorname{Re} \sum_{j=1}^J W_{j+1}^* S(z) \mathbb{M}(z) W_j = 0. \quad (25)$$

The sum of (24) and (25) yields:

$$W_1^* S(z) W_1 - W_{J+1}^* S(z) W_{J+1} + \sum_{j=1}^J \underbrace{W_j^* (\mathbb{M}(z)^* S(z) \mathbb{M}(z) - S(z)) W_j}_{\geq 0} = 0.$$

Letting J tend to infinity, we obtain $W_1^* S(z) W_1 \leq 0$ for all $W_1 \in \mathbb{E}^s(z)$. Using the properties of S , we obtain:

$$\forall z \in \mathcal{O} \cap \mathcal{U}, \quad \forall W_1 \in \mathbb{E}^s(z), \quad K |\pi^u W_1| \leq |\pi^s W_1|.$$

The end of the analysis follows [14]. Writing $\pi^s W_1 = W_1 - \pi^u W_1$, we get:

$$\forall z \in \mathcal{O} \cap \mathcal{U}, \quad \forall W_1 \in \mathbb{E}^s(z), \quad (K-1) |\pi^u W_1| \leq |W_1|. \quad (26)$$

The estimate (26) shows that the mapping:

$$\begin{aligned} \Phi(z) : \mathbb{E}^s(z) &\longrightarrow \underline{\mathbb{E}}^s \\ W_1 &\longmapsto \pi^s W_1, \end{aligned}$$

is injective, and is therefore an isomorphism because the dimensions of $\mathbb{E}^s(z)$ and $\underline{\mathbb{E}}^s$ are equal. We can then write the inverse mapping $\Phi(z)^{-1}$ in the following way:

$$\begin{aligned} \Phi(z)^{-1} : \underline{\mathbb{E}}^s &\longrightarrow \mathbb{E}^s(z) \\ \underline{W}_1 &\longmapsto \underline{W}_1 + \varphi(z) \underline{W}_1, \end{aligned}$$

where $\varphi(z)$ is a linear mapping from $\underline{\mathbb{E}}^s$ to $\underline{\mathbb{E}}^u$. Using (26) we have:

$$\forall z \in \mathcal{O} \cap \mathcal{U}, \quad \forall \underline{W}_1 \in \underline{\mathbb{E}}^s, \quad |\varphi(z) \underline{W}_1| \leq \frac{1}{K-2} |\underline{W}_1|. \quad (27)$$

Combining the fact that $\Phi(z)^{-1}$ is surjective and the estimate (27) shows that the space $\mathbb{E}^s(z)$ tends to $\underline{\mathbb{E}}^s$ as $z \in \mathcal{U}$ tends to \underline{z} . We have thus extended \mathbb{E}^s to $\overline{\mathcal{U}}$. Following again [14] and performing some similar calculations as above, we can show that this extension of \mathbb{E}^s to $\overline{\mathcal{U}}$ defines a continuous vector bundle over $\overline{\mathcal{U}}$. We refer to [14] for more details.

From now on, we further assume that for all $z \in \overline{\mathcal{U}}$ we have $\mathbb{E}^s(z) \cap \operatorname{Ker} B(z) = \{0\}$, which is the uniform Kreiss-Lopatinskii condition. We are going to show that the scheme (2) is strongly stable. We first consider the case of large frequencies z :

Lemma 4 ([9]). *There exist two constants $R_0 \geq 2$ and $C_0 \geq 0$ that depend only on the scheme (2) such that for all z verifying $|z| \geq R_0$, for all $(F_j) \in \ell^2$ and all $g_{-r+1}, \dots, g_0 \in \mathbb{C}^N$, the resolvent equation (5) has a unique solution $(w_j) \in \ell^2$ and this solution satisfies:*

$$\sum_{j \geq -r+1} |w_j|^2 \leq C_0 \left(\sum_{j \geq 1} |F_j|^2 + \sum_{j=-r+1}^0 |g_j|^2 \right).$$

Proof. For $z = \infty$, the resolvent equation (5) formally reduces to:

$$\begin{cases} w_j = F_j, & j \geq 1, \\ w_j - B_{j,-1} w_1 = g_j, & j = -r+1, \dots, 0. \end{cases} \quad (28)$$

We let $\mathcal{R}(F, g)$ denote the sequence $(w_j)_{j \geq -r+1}$ solution to (28). It is clear that \mathcal{R} is a continuous linear operator on the Hilbert space $\ell^2 \times \mathbb{C}^{Nr}$ with values in ℓ^2 . For $z \in \mathcal{U}$, the resolvent equation (5) may be recast as:

$$w = \mathcal{R} \left(F + \left(\sum_{\sigma=0}^s z^{-\sigma-1} Q_{\sigma} w_j \right)_{j \geq 1}, g + \left(\sum_{\sigma=0}^s z^{-\sigma-1} B_{j,\sigma} w_1 \right)_{j=-r+1, \dots, 0} \right),$$

which is a fixed point problem. This problem may be solved, for large enough $|z|$, by applying the Banach fixed point Theorem, which gives a unique solution $w \in \ell^2$. We omit the details that are almost straightforward. The estimate of the solution $w \in \ell^2$ in terms of F and g is also straightforward. \square

It remains to solve the resolvent equation when $|z| \in]1, R_0]$ and to obtain the corresponding estimate (6). We are first going to prove that for $|z| \in]1, R_0]$, the resolvent equation satisfies an a priori estimate. More precisely, let $(w_j) \in \ell^2$, and let $F \in \ell^2$, g_{-r+1}, \dots, g_0 be defined such that (5) holds. We rewrite (5) under the equivalent form (9), as detailed in the introduction. Then we have:

Lemma 5. *There exists a constant $c_0 > 0$ and there exists a \mathcal{C}^∞ function S on the closed annulus $\{1 \leq |\zeta| \leq R_0\}$ with values in $\mathcal{M}_{N(p+r)}(\mathbb{C})$ such that the following properties hold for all $z \in \{1 \leq |\zeta| \leq R_0\}$:*

- $S(z)$ is hermitian,
- $\mathbb{M}(z)^* S(z) \mathbb{M}(z) - S(z) \geq c_0 (|z| - 1)/|z| I$,
- for all $W \in \mathbb{C}^{N(p+r)}$, $W^* S(z) W \geq c_0 |W|^2 - |B(z) W|^2 / c_0$.

The proof of Lemma 5 follows from the existence of a K -symmetrizer and from the uniform Kreiss-Lopatinskii condition. We first construct the symmetrizer S in the neighborhood of any point of the annulus $\{1 \leq |\zeta| \leq R_0\}$ by choosing K sufficiently large. We refer to [15] for a similar analysis. The construction of a global smooth symmetrizer on the annulus requires a partition of unity as in the analysis of the continuous problem, see e.g. [2, 11]. Using the symmetrizer S of Lemma 5, we can perform some similar calculations to what we have done at the beginning of this section to obtain relations (24) and (25). Here the source term \widetilde{F}_j has to be taken into account, while it did not appear in (23). We obtain:

$$W_1^* S(z) W_1 + \sum_{j=1}^{+\infty} W_j^* (\mathbb{M}(z)^* S(z) \mathbb{M}(z) - S(z)) W_j = -\operatorname{Re} \sum_{j=1}^{+\infty} (S(z) W_{j+1} + S(z) \mathbb{M}(z) W_j)^* \widetilde{F}_j.$$

Using some uniform bounds for $S(z)$ and $\mathbb{M}(z)$ on the annulus as well as the properties of S given in Lemma 5, we end up with:

$$\frac{|z|-1}{|z|} \sum_{j \geq 1} |W_j|^2 + |W_1|^2 \leq C_1 \left\{ \frac{|z|}{|z|-1} \sum_{j \geq 1} |\widetilde{F_j}|^2 + |g|^2 \right\},$$

for some appropriate numerical constant C_1 . This estimate immediately yields the a priori estimate (6) for (5) when $|z| \in]1, R_0]$ by using the definition of $(W_j), (\widetilde{F_j}), g$.

Up to now, we have only proved an a priori estimate for the resolvent equation (5) when $|z| \in]1, R_0]$. It remains to show that (5) may be solved in a unique way for arbitrary source terms. This step of the proof did not appear in [9]. We make use of the following general result:

Lemma 6. *Let E be a Banach space, and let \mathcal{T} denote a nonempty connected set. Let L be a continuous function defined on \mathcal{T} with values in the space $\mathcal{L}(E)$ of continuous linear maps on E . Assume moreover that the two following conditions are satisfied:*

- *there exists a constant $C_0 > 0$ such that for all $t \in \mathcal{T}$ and for all $x \in E$, we have $|x|_E \leq C_0 |L(t)x|_E$,*
- *there exists some $t_0 \in \mathcal{T}$ such that $L(t_0)$ is an isomorphism.*

Then $L(t)$ is an isomorphism for all $t \in \mathcal{T}$.

The proof of Lemma 6 is given in appendix B. Lemma 6 shows that the resolvent equation can be uniquely solved for all $z \in \mathcal{U}$. Indeed, for all $z \in \mathcal{U}$, we define the mapping:

$$L(z) : \ell^2 \longrightarrow \ell^2$$

$$w \longmapsto L(z)w \quad \text{with} \quad (L(z)w)_j := \begin{cases} w_j - \sum_{\sigma=0}^s z^{-\sigma-1} Q_\sigma w_j, & j \geq 1, \\ w_j - \sum_{\sigma=-1}^s z^{-\sigma-1} B_{j,\sigma} w_1, & j = -r+1, \dots, 0. \end{cases}$$

We can easily verify that L is a continuous map from \mathcal{U} to the set $\mathcal{L}(\ell^2)$ and for all $|z| \geq R_0$, $L(z)$ is an isomorphism thanks to Lemma 4. We can then apply Lemma 6 on every annulus $\{1+\varepsilon \leq |z| \leq 1/\varepsilon\}$, $\varepsilon > 0$ small enough, and conclude that $L(z)$ is an isomorphism for all $z \in \mathcal{U}$. This shows that the resolvent equation (5) can be uniquely solved for all $z \in \mathcal{U}$ and all source terms $F \in \ell^2$, $g_{-r+1}, \dots, g_0 \in \mathbb{C}^N$. The corresponding estimate of the unique solution $w \in \ell^2$ follows from the analysis above and we have therefore proved that the scheme (2) is strongly stable. The proof of Theorem 2 is complete.

5 Proof of Theorem 3

Using the result of Lemma 1, we know that the “stable” subspace $\mathbb{E}^s(z)$ of $\mathbb{M}(z)$ has constant dimension Nr for all $z \in \mathcal{U}$ and varies holomorphically with respect to $z \in \mathcal{U}$. Using Theorem 1 we also know that \mathbb{M} satisfies the discrete block structure condition. We are now going to build a K -symmetrizer in the neighborhood of any point $\underline{z} \in \mathcal{U}$.

We first consider a point $\underline{z} \in \mathcal{U}$. Using Theorem 1 we can reduce the matrix $\mathbb{M}(\underline{z})$ to a block-diagonal form⁵:

$$\underline{T}^{-1} \mathbb{M}(\underline{z}) \underline{T} = \text{diag} (\underline{\mathbb{M}}^s, \underline{\mathbb{M}}^u).$$

⁵Recall that for $\underline{z} \in \mathcal{U}$, blocks of the third and fourth type do not appear in the discrete block structure.

The first block $\underline{\mathbb{M}}^s$ satisfies $(\underline{\mathbb{M}}^s)^* \underline{\mathbb{M}}^s \leq (1 - \delta) I$ and the second block $\underline{\mathbb{M}}^u$ satisfies $(\underline{\mathbb{M}}^u)^* \underline{\mathbb{M}}^u \geq (1 + \delta) I$ for some appropriate constant $\delta > 0$. It is then easy to construct a K -symmetrizer, the spaces $\underline{\mathbb{E}}^s, \underline{\mathbb{E}}^u$ in the decomposition of $\mathbb{C}^{N(p+r)}$ being nothing but the “stable” and “unstable” subspace of $\mathbb{M}(\underline{z})$ associated with eigenvalues in \mathbb{D} and \mathcal{U} . We let $\underline{\pi}^s, \underline{\pi}^u$ denote the corresponding projectors. Then there exists a constant $c_0 > 0$ such that for all $W \in \mathbb{C}^{N(p+r)}$, the following estimates hold:

$$c_0 |\underline{\pi}^s W| \leq |(\underline{T}^{-1} W)_1| \leq \frac{1}{c_0} |\underline{\pi}^s W|, \quad c_0 |\underline{\pi}^u W| \leq |(\underline{T}^{-1} W)_2| \leq \frac{1}{c_0} |\underline{\pi}^u W|,$$

where we use the notation $U = (U_1, U_2)$ with $U_1 \in \mathbb{C}^{Nr}, U_2 \in \mathbb{C}^{Np}$ for any vector $U \in \mathbb{C}^{N(p+r)}$. The symmetrizer S can be chosen independent of z as follows:

$$S := (\underline{T}^{-1})^* \text{diag} \left(-\frac{c_0^2}{2} I, \frac{K^2 + 1/2}{c_0^2} I \right) \underline{T}^{-1}.$$

Then $\mathbb{M}(\underline{z})^* S \mathbb{M}(\underline{z}) - S$ is positive definite, and we have:

$$W^* S W \geq \left(K^2 + \frac{1}{2} \right) |\underline{\pi}^u W|^2 - \frac{1}{2} |\underline{\pi}^s W|^2.$$

Choosing S independent of z , and z sufficiently close to \underline{z} , we have constructed a K -symmetrizer near \underline{z} .

We turn to the case $\underline{z} \in \mathbb{S}^1$. Using Theorem 1, we can first reduce \mathbb{M} to a block-diagonal form with a holomorphic change of basis on a neighborhood \mathcal{O} of \underline{z} :

$$T(z)^{-1} \mathbb{M}(z) T(z) = \text{diag} \left(\mathbb{M}^s(z), \mathbb{M}^u(z), m_1^s(z), \dots, m_l^s(z), m_1^u(z), \dots, m_j^u(z), H_1(z), \dots, H_K(z), P_1(z), \dots, P_L(z) \right),$$

where in the terminology of definition 2 \mathbb{M}^s is a block of the first type, \mathbb{M}^u is a block of the second type, the m_i^s 's (resp. m_j^u 's) are blocks of the third type with $\underline{z}(m_i^s)'(\underline{z}) \overline{m_i^s(\underline{z})} < 0$ (resp. $\underline{z}(m_j^u)'(\underline{z}) \overline{m_j^u(\underline{z})} > 0$). Eventually the blocks H_k 's and P_ℓ 's are of the fourth type, the size of each block P_ℓ is 2, and we have:

$$\begin{aligned} \det(H_k(z) - \kappa I) &= \vartheta(\kappa, z) (z - \Lambda_k(\kappa)), \quad \Lambda_k(\underline{\kappa}_k) = \underline{z}, \quad \Lambda_k'(\underline{\kappa}_k) = 0, \\ \det(P_\ell(z) - \kappa I) &= \vartheta(\kappa, z) (z - \lambda_\ell(\kappa)), \quad \lambda_\ell(\underline{\kappa}_\ell) = \underline{z}, \quad \lambda_\ell'(\underline{\kappa}_\ell) = 0, \quad \text{Re}(\underline{\kappa}_\ell^2 \underline{\lambda}_\ell''(\underline{\kappa}_\ell)) > 0. \end{aligned}$$

In the relations above, the numbers $\underline{\kappa}_1, \dots, \underline{\kappa}_K, \underline{\kappa}_1, \dots, \underline{\kappa}_L$ all belong to \mathbb{S}^1 . Moreover, we have $\Lambda_k(\kappa) \in \mathbb{S}^1$ for all $\kappa \in \mathbb{S}^1$ close enough to $\underline{\kappa}_k$. The generic notation ϑ is again used to denote a holomorphic function of its arguments that does not vanish. The blocks H_k 's will be referred to as “hyperbolic blocks” while the blocks P_ℓ 's will be referred to as “parabolic blocks”. This terminology will be clarified below.

To prove the existence of a K -symmetrizer for \mathbb{M} in the neighborhood of \underline{z} , we are first going to construct a K -symmetrizer for the block-diagonal matrix $\tilde{\mathbb{M}} := T^{-1} \mathbb{M} T$. The decomposition of $\mathbb{C}^{N(p+r)}$ is defined as follows. For each block H_k , we let μ_k denote the number of eigenvalues of $H_k(z)$ in \mathbb{D} for $z \in \mathcal{O} \cap \mathcal{U}$. This number is independent of z thanks to Lemma 1. We also let ν_k (resp. ν_s, ν_u) denote the size of the block H_k (resp. $\mathbb{M}^s, \mathbb{M}^u$). Then we define:

$$\begin{aligned} \tilde{\underline{\mathbb{E}}}^s &:= \mathbb{C}^{\nu_s} \oplus \{0\} \oplus_{i=1}^I \mathbb{C}^1 \oplus_{j=1}^J \{0\} \oplus_{k=1}^K \mathbb{C}^{\mu_k} \times \{0\} \oplus_{\ell=1}^L \mathbb{C}^1 \times \{0\}, \\ \tilde{\underline{\mathbb{E}}}^u &:= \{0\} \oplus \mathbb{C}^{\nu_u} \oplus_{i=1}^I \{0\} \oplus_{j=1}^J \mathbb{C}^1 \oplus_{k=1}^K \{0\} \times \mathbb{C}^{\nu_k - \mu_k} \oplus_{\ell=1}^L \{0\} \times \mathbb{C}^1, \end{aligned}$$

where the decomposition of $\tilde{\mathbb{E}}^s, \tilde{\mathbb{E}}^u$ follows the block-diagonal form of $\tilde{\mathbb{M}}(z)$, and where the notation $\mathbb{C}^{\mu_k} \times \{0\}$ (resp. $\{0\} \times \mathbb{C}^{\nu_k - \mu_k}$) denotes the subspace of \mathbb{C}^{ν_k} spanned by the μ_k first (resp. $\nu_k - \mu_k$ last) vectors of the canonical basis. Observe that each block $P_l(z)$ has exactly one eigenvalue in \mathbb{D} and one eigenvalue in \mathcal{U} for $z \in \mathcal{O} \cap \mathcal{U}$, so the dimension of $\tilde{\mathbb{E}}^s$ (resp. $\tilde{\mathbb{E}}^u$) equals Nr (resp. Np). We let $\tilde{\pi}^s, \tilde{\pi}^u$ denote the projectors associated with the decomposition $\mathbb{C}^{N(p+r)} = \tilde{\mathbb{E}}^s \oplus \tilde{\mathbb{E}}^u$.

We now fix a constant $K \geq 1$. The symmetrizer $\tilde{S}(z)$ for the matrix $\tilde{\mathbb{M}}(z)$ is chosen in block-diagonal form:

$$\tilde{S}(z) := \text{diag}(\mathbf{S}^s, \mathbf{S}^u, r_1^s, \dots, r_I^s, r_1^u, \dots, r_J^u, R_1(z), \dots, R_K(z), S_1(z), \dots, S_L(z)),$$

and we detail now the construction of each block in the symmetrizer.

- Construction of the constant matrices \mathbf{S}^s and \mathbf{S}^u . These matrices are defined as in the case $\underline{z} \in \mathcal{U}$ by $\mathbf{S}^s := -I/2$ and $\mathbf{S}^u := (K^2 + 1/2)I$. For $z \in \overline{\mathcal{U}}$ sufficiently close to \underline{z} , we have:

$$\mathbb{M}^s(z)^* \mathbf{S}^s \mathbb{M}^s(z) - \mathbf{S}^s \geq c \frac{|z| - 1}{|z|} I, \quad \mathbb{M}^u(z)^* \mathbf{S}^u \mathbb{M}^u(z) - \mathbf{S}^u \geq c \frac{|z| - 1}{|z|} I,$$

for some appropriate constant $c > 0$.

- Construction of the numbers $r_1^s, \dots, r_I^s, r_1^u, \dots, r_J^u$. Let us begin with the following:

Lemma 7. *We have $m_1^s(z), \dots, m_I^s(z) \in \mathbb{D}$ and $m_1^u(z), \dots, m_J^u(z) \in \mathcal{U}$ for all $z \in \mathcal{O} \cap \mathcal{U}$. Moreover up to shrinking the neighborhood \mathcal{O} of \underline{z} , there exists a constant $c > 0$ such that the following inequalities hold for all $z \in \mathcal{O} \cap \overline{\mathcal{U}}$:*

$$\begin{aligned} \forall i = 1, \dots, I, \quad |m_i^s(z)|^2 - 1 &\leq -c \frac{|z| - 1}{|z|}, \\ \forall j = 1, \dots, J, \quad |m_j^u(z)|^2 - 1 &\geq c \frac{|z| - 1}{|z|}. \end{aligned}$$

Proof. There is no loss of generality in assuming that \mathcal{O} is connected. Then for $z \in \mathcal{O} \cap \mathcal{U}$, $m_i^s(z)$ and $m_j^u(z)$ are eigenvalues of $\mathbb{M}(z)$ so they cannot belong to \mathbb{S}^1 . Performing a Taylor expansion of $m_i^s((1 + \varepsilon)\underline{z})$ and $m_j^u((1 + \varepsilon)\underline{z})$ for $\varepsilon > 0$ small, we see that $m_i^s((1 + \varepsilon)\underline{z}) \in \mathbb{D}$ and $m_j^u((1 + \varepsilon)\underline{z}) \in \mathcal{U}$ for $\varepsilon > 0$ small enough. Since $\mathcal{O} \cap \mathcal{U}$ is connected, we have $m_i^s(z) \in \mathbb{D}$ and $m_j^u(z) \in \mathcal{U}$ for all $z \in \mathcal{O} \cap \mathcal{U}$.

For simplicity, let us now deal with m_1^s . For τ in a small neighborhood of 0, we define:

$$h(\tau) := \ln \frac{m_1^s(\underline{z} e^\tau)}{m_1^s(\underline{z})},$$

where \ln denotes the standard logarithm defined on $\mathbb{C} \setminus \mathbb{R}^-$. We have $h'(0) = \underline{z} (m_1^s)'(\underline{z}) \overline{m_1^s(\underline{z})} < 0$, and we also know that $h(\tau)$ has negative real part when τ has positive real part. In particular, $h(\tau)$ has nonpositive real part when τ is purely imaginary. Then a direct Taylor expansion yields the estimate⁶:

$$\text{Re } h(\tau) \leq \frac{h'(0)}{4} \text{Re } \tau,$$

for $\text{Re } \tau \geq 0$ and τ close to 0. The estimate for $|m_1^s(z)|^2$ easily follows. The case of the m_j^u 's is similar. \square

⁶Here we use the crucial fact that $\text{Re } h(iy) \leq 0$ for all real y close to 0. The only inequality $h'(0) < 0$ is not sufficient to yield the estimate. This part of the analysis was not so clear in [9, page 678].

For all $i = 1, \dots, I$ we define $r_i^s := -1/2$ and for all $j = 1, \dots, J$ we define $r_j^s := K^2 + 1/2$. Using Lemma 7, we have:

$$r_i^s (|m_i^s(z)|^2 - 1) \geq c \frac{|z| - 1}{|z|}, \quad r_j^u (|m_j^u(z)|^2 - 1) \geq c \frac{|z| - 1}{|z|},$$

for some appropriate constant $c > 0$ and for all $z \in \mathcal{O} \cap \overline{\mathcal{U}}$.

• Construction of the symmetrizers $S_1(z), \dots, S_L(z)$. This is the main difficulty in the proof of Theorem 3 where new arguments are needed. For simplicity, we are going to construct the symmetrizer $S_1(z)$ associated with the block $P_1(z)$. Recall that we have:

$$\det(P_1(z) - \kappa I) = \vartheta(\kappa, z) (z - \lambda_1(\kappa)), \quad \lambda_1(\kappa_1) = \underline{z}, \quad \lambda_1'(\kappa_1) = 0, \quad \operatorname{Re}(\kappa_1^2 \bar{\underline{z}} \lambda_1''(\kappa_1)) > 0.$$

According to Theorem 1, we even know that the lower left coefficient ω_1 of $P_1'(\underline{z})$ satisfies $\omega_1 \neq 0$ and $\operatorname{Re}(\bar{\kappa}_1 \omega_1 \underline{z}) \geq 0$. Indeed these two conditions can be seen to be equivalent to the property stated in definition 2 for a 2×2 block. We are first going to prove that the real part of $\bar{\kappa}_1 \omega_1 \underline{z}$ is positive. To prove this property, let us write:

$$P_1(z) = \kappa_1 \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} p_1(z) & p_3(z) \\ p_2(z) & p_4(z) \end{pmatrix}, \quad p_2'(\underline{z}) = \omega_1.$$

Using the fact that κ is an eigenvalue of $P_1(\lambda_1(\kappa))$ for all κ close to κ_1 , we obtain:

$$(\kappa - \kappa_1)^2 = (\kappa - \kappa_1) (p_1(z) + p_4(z)) - p_1(z) p_4(z) + p_2(z) (\kappa_1 + p_3(z))|_{z=\lambda_1(\kappa)}.$$

Differentiating twice and evaluating at $\kappa = \kappa_1$, we obtain $\omega_1 = 2/(\kappa_1 \lambda_1''(\kappa_1))$. The inequality $\operatorname{Re}(\bar{\kappa}_1 \omega_1 \underline{z}) > 0$ immediately follows from the assumption on λ_1 .

To construct a symmetrizer for P_1 , it is easier to work in “flat” coordinates. In the original coordinates, z belongs to the exterior of a disc so the curvature of the disc makes the calculations heavier. We shall reduce below to coordinates in a half-space. Let us denote ξ_1 the argument of κ_1 , that is $\kappa_1 = \exp(i \xi_1)$. Then we can define a matrix that depends holomorphically on z in a neighborhood of \underline{z} by the following formula:

$$\mathbf{P}_1(z) := i \xi_1 I + \sum_{n=1}^{+\infty} \frac{(-1)^{n-1}}{n} (\bar{\kappa}_1 P_1(z) - I)^n.$$

This matrix satisfies $P_1(z) = \exp \mathbf{P}_1(z)$ for all z close enough to \underline{z} . We compute:

$$\mathbf{P}_1(\underline{z}) = \begin{pmatrix} i \xi_1 & 1 \\ 0 & i \xi_1 \end{pmatrix} = i \xi_1 I + N. \quad (29)$$

The first derivative $\mathbf{P}_1'(\underline{z})$ is related to $P_1'(\underline{z})$ through the differential of the exponential function at $\mathbf{P}_1(\underline{z})$. More precisely, we recall the following formula that is fully justified in [17, page 78]:

$$\mathrm{d} \exp|_A B := \frac{\mathrm{d}}{\mathrm{d} z} \exp(A + z B)|_{z=0} = \exp(A) \sum_{p=0}^{+\infty} \frac{(-1)^p}{(p+1)!} (\mathrm{ad} A)^p B, \quad (\mathrm{ad} A) B := A B - B A. \quad (30)$$

We differentiate the relation $P_1(z) = \exp \mathbf{P}_1(z)$, evaluate at $z = \underline{z}$ and use (29), (30) to obtain:

$$\bar{\kappa}_1 P_1'(\underline{z}) = \mathbf{P}_1'(\underline{z}) + \frac{1}{2} (N \mathbf{P}_1'(\underline{z}) + \mathbf{P}_1'(\underline{z}) N) + \frac{1}{6} N \mathbf{P}_1'(\underline{z}) N. \quad (31)$$

The nilpotent matrix N is defined in (29). The relation (31) shows that the lower left element of $\mathbf{P}'_1(\underline{z})$ equals $\overline{\kappa_1} \omega_1$. Eventually, we define the matrix:

$$\mathbb{P}_1(\tau) := Q_0^{-1} \mathbf{P}_1(\underline{z} e^\tau) Q_0, \quad Q_0 := \text{diag}(-i, 1). \quad (32)$$

It depends holomorphically on τ in a neighborhood of 0 and satisfies $\mathbb{P}_1(0) = i(\xi_1 I + N)$. Moreover, the lower left coefficient of $\mathbb{P}'_1(0)$, that is from now on denoted α_1 , has negative imaginary part.

If we want to see the analogy and the differences with the analysis of the continuous problem, for which we refer to [11, 2], we need to introduce a final change of basis that is due to Ralston (see [19] for the relevance of this change of basis):

$$Q(\tau) := (-i(\mathbb{P}_1(\tau) - i\xi_1 I) e_2 \quad e_2), \quad e_2 := \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

The matrix $Q(\tau)$ depends holomorphically on τ , it attains the value I for $\tau = 0$ and is therefore invertible for τ close to 0. Moreover, the matrix $Q^{-1} \mathbb{P}_1 Q$ reads:

$$\check{\mathbb{P}}_1(\tau) := Q(\tau)^{-1} \mathbb{P}_1(\tau) Q(\tau) = i(\xi_1 I + N) + \begin{pmatrix} b_1(\tau) & 0 \\ b_2(\tau) & 0 \end{pmatrix}, \quad (33)$$

where the functions b_1, b_2 are holomorphic and vanish for $\tau = 0$. Moreover, we compute $b'_2(0) = \alpha_1$. Let us recall that α_1 denotes the lower left coefficient of $\mathbb{P}'_1(0)$ and has negative imaginary part.

After all these preliminary transformations, the construction of the symmetrizer S_1 relies on the following:

Lemma 8. *There exist a neighborhood of 0, a \mathcal{C}^∞ mapping \check{S} defined on this neighborhood with values in $\mathcal{M}_2(\mathbb{C})$, and there exists a constant $c > 0$ such that for all τ close to 0 the following properties hold:*

- $\check{S}(\tau)$ is hermitian,
- $\text{Re}(\check{S}(\tau) \check{\mathbb{P}}_1(\tau)) \geq c(\text{Re } \tau + (\text{Im } \tau)^2) I$ if $\text{Re } \tau \geq 0$,
- for all $W \in \mathbb{C}^2$, $W^* \check{S}(0) W \geq (K^2 + 1/2) |W_2|^2 - |W_1|^2/2$.

Proof. The proof follows the ideas in [11] but we can not adopt the same construction here. Some new ingredients are needed⁷. The additional term $(\text{Im } \tau)^2$ in the estimate explains the terminology “parabolic block” introduced above for P_1 . This new term appears because of the “dissipation” assumption made on λ_1 .

In all the proof of Lemma 8, we use the notation $\tau = \gamma + i\delta$. The symmetrizer \check{S} is sought, as in [11], under the form:

$$\check{S}(\tau) = \check{S}_0(\delta) + \gamma \check{S}_1,$$

where \check{S}_0 is a \mathcal{C}^∞ function of the real variable δ , and \check{S}_1 is a constant hermitian matrix⁸. The matrix $\check{S}_0(\delta)$ is chosen under the form:

$$\check{S}_0(\delta) := \begin{pmatrix} 0 & \overline{\alpha_1} + c_0 \delta \\ \alpha_1 + c_0 \delta & a_0 \end{pmatrix},$$

⁷The problem treated by Kreiss [11] in the continuous case corresponds to $\alpha_1 \in \mathbb{R} \setminus \{0\}$ while here, the “dissipation” assumption on λ_1 yields $\text{Im } \alpha_1 < 0$.

⁸In [11], $\check{S}_0(\delta)$ can be chosen with real coefficients and such that the real part of $\check{S}_0(\delta) \check{\mathbb{P}}_1(i\delta)$ is identically zero. In our case, it is not possible to choose a matrix $\check{S}_0(\delta)$ with real coefficients so that the real part of $\check{S}_0(\delta) \check{\mathbb{P}}_1(i\delta)$ is nonnegative.

where a_0, c_0 are real constants. Decomposing $b_2(i\delta) = i\alpha_1\delta + \delta^2 b_3(\delta)$ for some analytic function b_3 , (33) and our choice of \check{S}_0 yield:

$$\begin{aligned} \operatorname{Re}(\check{S}_0(\delta)\check{\mathbb{P}}_1(i\delta)) &= \begin{pmatrix} c_0\delta^2(-\operatorname{Im}\alpha_1 + \delta b_3(\delta)) + \delta^2 \operatorname{Re}(\overline{\alpha_1}b_3(\delta)) & \star \\ ((\alpha_1 + c_0\delta)b_1(i\delta) + a_0b_2(i\delta))/2 & -\operatorname{Im}\alpha_1 \end{pmatrix} \\ &= \begin{pmatrix} \delta^2(-c_0\operatorname{Im}\alpha_1 + \operatorname{Re}(\overline{\alpha_1}b_3(0))) & \star \\ i\alpha_1\delta(a_0 + b'_1(0))/2 & -\operatorname{Im}\alpha_1 \end{pmatrix} + \begin{pmatrix} O(\delta^3) & \star \\ O(\delta^2) & 0 \end{pmatrix}, \end{aligned}$$

where the \star coefficients are such that the matrices above are hermitian. We first fix the constant a_0 large enough such that the following estimate holds:

$$\forall W \in \mathbb{C}^2, \quad W^* \check{S}_0(0)W \geq (K^2 + 1/2)|W_2|^2 - |W_1|^2/2.$$

Once a_0 is fixed, we can choose c_0 large enough such that for all δ sufficiently small, we have:

$$\operatorname{Re}(\check{S}_0(\delta)\check{\mathbb{P}}_1(i\delta)) \geq \delta^2 I.$$

Defining the matrix \check{S}_1 as in [11] by:

$$\check{S}_1 := \begin{pmatrix} 0 & ig \\ -ig & 0 \end{pmatrix},$$

we compute:

$$\begin{aligned} \operatorname{Re}(\check{S}(\tau)\check{\mathbb{P}}_1(\tau)) &= \operatorname{Re}(\check{S}_0(\delta)\check{\mathbb{P}}_1(i\delta)) + \gamma \operatorname{Re}(\check{S}_0(0)\check{\mathbb{P}}'_1(0) + \check{S}_1\check{\mathbb{P}}_1(0)) + \gamma o(1), \\ \operatorname{Re}(\check{S}_0(0)\check{\mathbb{P}}'_1(0) + \check{S}_1\check{\mathbb{P}}_1(0)) &= \begin{pmatrix} |\alpha_1|^2 & \star \\ \alpha_1(a_0 + b'_1(0))/2 & g \end{pmatrix}. \end{aligned}$$

Choosing g large enough, we obtain all the properties stated in Lemma 8 for the symmetrizer $\check{S}(\tau)$. \square

Using the symmetrizer \check{S} of Lemma 8, we define $S_1(z) := (Q(\tau)^{-1}Q_0^{-1})^*\check{S}(\tau)Q(\tau)^{-1}Q_0^{-1}$ where $\tau = \ln(z/\underline{z})$. The matrix $S_1(z)$ is hermitian and S_1 is a \mathcal{C}^∞ function of z . Using $Q(0) = I$ and the explicit expression of Q_0 , see (32), we also obtain:

$$\forall W \in \mathbb{C}^2, \quad W^* S_1(\underline{z})W \geq (K^2 + 1/2)|W_2|^2 - |W_1|^2/2.$$

It remains to check that $S_1(z)$ symmetrizes $P_1(z)$. This is done by following the calculations of [9, page 685] so we omit the details. With some appropriate constant $c > 0$, we obtain:

$$P_1(z)^* S_1(z) P_1(z) - S_1(z) \geq c \frac{|z| - 1}{|z|} I,$$

for all $z \in \mathcal{U}$ close enough to \underline{z} .

• Construction of $R_1(z), \dots, R_K(z)$. Following the analysis of the preceeding case, we can first write each hyperbolic block $H_k(z)$ under the form $H_k(z) = \exp \mathbf{H}_k(z)$. Then up to a holomorphic change of basis Q , we can reduce \mathbf{H}_k to the following form (see [19]):

$$Q(\tau)^{-1} \mathbf{H}_k(\underline{z}e^\tau) Q(\tau) = \begin{pmatrix} i\underline{\xi}_k & i & 0 & 0 \\ 0 & \ddots & \ddots & 0 \\ \vdots & & \ddots & i \\ 0 & \cdots & 0 & i\underline{\xi}_k \end{pmatrix} + \begin{pmatrix} b_1(\tau) & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ b_{\nu_k}(\tau) & 0 & \cdots & 0 \end{pmatrix}.$$

This relation holds for all τ in a sufficiently small neighborhood of 0. Moreover, following [13], we can show that $b'_{\nu_k}(0)$ is a nonzero real number and that the b_j 's are purely imaginary valued when τ is purely imaginary. The change of basis Q can be chosen such that $Q(0)$ is upper triangular. Then the analysis of [11], see also [2], shows that there exists a K -symmetrizer $R_k(z)$ for $\mathbf{H}_k(z)$. In particular, R_k satisfies:

$$W^* R_k(\underline{z}) W \geq (K^2 + 1/2) |W_2|^2 - 1/2 |W_1|^2,$$

where $W = (W_1, W_2)$, $W_1 \in \mathbb{C}^{\mu_k}$, $W_2 \in \mathbb{C}^{\nu_k - \mu_k}$. The calculations of [9, page 685] show that R_k is also a K -symmetrizer for the hyperbolic block $H_k = \exp \mathbf{H}_k$.

• End of the analysis. The definition of our symmetrizer $\tilde{S}(z)$ shows that \tilde{S} is a \mathcal{C}^∞ function of z and it symmetrizes $\tilde{\mathbb{M}} = T^{-1} M T$ in the following sense:

$$\tilde{\mathbb{M}}(z)^* \tilde{S}(z) \tilde{\mathbb{M}}(z) - \tilde{S}(z) \geq c \frac{|z| - 1}{|z|} I,$$

for $z \in \overline{\mathcal{W}}$ sufficiently close to \underline{z} . Moreover, the matrix $\tilde{S}(\underline{z})$ satisfies:

$$W^* \tilde{S}(\underline{z}) W \geq (K^2 + 1/2) |\underline{\pi}^u W|^2 - 1/2 |\underline{\pi}^s W|^2.$$

We now obtain the existence of a K -symmetrizer for \mathbb{M} by setting $\underline{\mathbb{E}}^{s,u} := T(\underline{z}) \tilde{\mathbb{E}}^{s,u}$, $S(z) := (T(z)^{-1})^* \tilde{S}(z) T(z)^{-1}$ and so on. The proof of Theorem 3 is complete.

In appendix C we extend our analysis of parabolic blocks to the case of an eigenvalue λ_j that satisfies a dissipation estimate⁹ of order 4, namely:

$$|\lambda_j(\underline{\kappa} e^{i\xi})| \leq 1 - c \xi^4, \quad c > 0, \quad (34)$$

for $\xi \in \mathbb{R}$ close to 0. Such an estimate may occur in various cases. For instance, one can have $\lambda'_j(\underline{\kappa}) \neq 0$. The symmetrizer construction in this case is treated in Theorem 3. One can also have the estimate (34) with $\lambda'_j(\underline{\kappa}) = 0$, $\lambda''_j(\underline{\kappa}) \neq 0$ and $\operatorname{Re}(\underline{\kappa}^2 \overline{\lambda_j(\underline{\kappa})} \lambda''_j(\underline{\kappa})) = 0$. This case is not covered by Theorem 3. As a matter of fact, this case corresponds to the limit where the strict inequality of condition *iii*) of Theorem 3 becomes an equality. We show in appendix C how to modify the symmetrizer construction in order to take this weaker assumption into account. Another possible case where (34) is valid occurs when $\lambda'_j(\underline{\kappa}) = \lambda''_j(\underline{\kappa}) = \lambda'''_j(\underline{\kappa}) = 0$ and $\lambda^{(4)}_j(\underline{\kappa}) \neq 0$. This situation is not covered by Theorem 3 and leads to a block of size 4. We shall give an example of numerical scheme where this highly degenerate situation occurs and detail the symmetrizer construction. The last possibility for (34) to occur is $\lambda'_j(\underline{\kappa}) = \lambda''_j(\underline{\kappa}) = 0$ and $\lambda'''_j(\underline{\kappa}) \neq 0$, which leads to a block of size 3.

We believe that the construction of a K -symmetrizer can be performed as long as the eigenvalues λ_j satisfy a dissipation estimate. This generalization is postponed to a future work.

6 Some old and new examples

In this section, we give some examples of numerical schemes for which our analysis applies. We do not focus on the numerical treatment of the boundary conditions but rather on the possible discretizations of the hyperbolic operator. For all the schemes below, the matrices $A_{l,\sigma}$ in (3)

⁹The condition *iii*) in Theorem 3 corresponds to a dissipation estimate of order 2, which is the simplest dissipation estimate one can get when $\lambda'_j(\underline{\kappa}) = 0$.

are polynomials of the matrix λA . All these matrices are therefore diagonalizable in a fixed basis where A is diagonal. Without loss of generality, we thus restrict in this section to the case of a scalar equation:

$$\partial_t u + a \partial_x u = f.$$

For simplicity, we assume $a > 0$ but the case $a < 0$ produces similar results. We recall that $\mathcal{A}(\kappa)$ denotes the symbol defined by (4). For scalar equations and one-step schemes ($s = 0$), this symbol is a complex number so the uniform power-boundedness and geometric regularity of eigenvalues reduce to the inequality $|\mathcal{A}(\kappa)| \leq 1$ for all $\kappa \in \mathbb{S}^1$.

Let us begin with some classical first and second order schemes:

The upwind scheme The scheme reads:

$$u_j^{n+1} = \lambda a u_{j-1}^n + (1 - \lambda a) u_j^n,$$

and we compute:

$$\mathcal{A}(\kappa) = \lambda a \kappa^{-1} + (1 - \lambda a).$$

The scheme is ℓ^2 -stable (and geometrically regular) if and only if $\lambda a \leq 1$. Assumption 1 is also satisfied if $\lambda a \leq 1$. If $\lambda a = 1$ we have $|\mathcal{A}(\kappa)| = 1$ for all $\kappa \in \mathbb{S}^1$ and $\mathcal{A}'(\kappa) = -\kappa^{-2} \neq 0$. If $\lambda a < 1$ we have $|\mathcal{A}(\kappa)| = 1$ if and only if $\kappa = 1$, and $\mathcal{A}'(1) = -\lambda a \neq 0$. The upwind scheme enters the framework of Theorem 3 if $\lambda a \leq 1$.

The Lax-Friedrichs scheme The scheme reads:

$$u_j^{n+1} = \frac{u_{j-1}^n + u_{j+1}^n}{2} - \frac{\lambda a}{2} (u_{j+1}^n - u_{j-1}^n),$$

and we compute:

$$\mathcal{A}(\kappa) = \frac{\kappa^{-1} + \kappa}{2} - \frac{\lambda a}{2} (\kappa - \kappa^{-1}).$$

The scheme is ℓ^2 -stable (and geometrically regular) if and only if $\lambda a \leq 1$. When $\lambda a = 1$ the scheme reduces to the upwind scheme analyzed above so we further assume $\lambda a < 1$. In that case, assumption 1 is satisfied and we have $\mathcal{A}(\kappa) \in \mathbb{S}^1$ if and only if $\kappa = \pm 1$. Moreover $\mathcal{A}'(\pm 1) = -\lambda a \neq 0$. If $\lambda a \leq 1$, the Lax-Friedrichs scheme enters the framework of Theorem 3.

The Lax-Wendroff scheme The scheme reads:

$$u_j^{n+1} = u_j^n - \frac{\lambda a}{2} (u_{j+1}^n - u_{j-1}^n) + \frac{\lambda^2 a^2}{2} (u_{j+1}^n + u_{j-1}^n - 2u_j^n),$$

and we compute:

$$\mathcal{A}(\kappa) = 1 - \frac{\lambda a}{2} (\kappa - \kappa^{-1}) + \frac{\lambda^2 a^2}{2} (\kappa + \kappa^{-1} - 2).$$

The scheme is ℓ^2 -stable (and geometrically regular) if and only if $\lambda a \leq 1$. When $\lambda a = 1$ the scheme reduces to the upwind scheme so we further assume $\lambda a < 1$. Assumption 1 is satisfied if $\lambda a < 1$. In that case, we have $\mathcal{A}(\kappa) \in \mathbb{S}^1$ if and only if $\kappa = 1$, and $\mathcal{A}'(1) = -\lambda a \neq 0$. If $\lambda a \leq 1$, the Lax-Wendroff scheme enters the framework of Theorem 3.

The leap-frog scheme The scheme reads:

$$u_j^{n+1} = u_j^{n-1} - \lambda a (u_{j+1}^n - u_{j-1}^n),$$

and we compute:

$$\mathcal{A}(\kappa) = \begin{pmatrix} -\lambda a (\kappa - \kappa^{-1}) & 1 \\ 1 & 0 \end{pmatrix}.$$

The scheme is ℓ^2 -stable and the eigenvalues of $\mathcal{A}(\kappa)$ are geometrically regular if and only if $\lambda a < 1$. In that case, assumption 1 is satisfied. The eigenvalues of $\mathcal{A}(e^{i\xi})$ are:

$$z_1(\xi) = -i \lambda a \sin \xi + \sqrt{1 - \lambda^2 a^2 \sin^2 \xi}, \quad z_2(\xi) = -i \lambda a \sin \xi - \sqrt{1 - \lambda^2 a^2 \sin^2 \xi}.$$

We have $z_1(\xi), z_2(\xi) \in \mathbb{S}^1$ for all $\xi \in \mathbb{R}$. Since $\lambda a < 1$, we can also construct eigenvectors $E_1(\xi)$ and $E_2(\xi)$ associated with the eigenvalues $z_1(\xi), z_2(\xi)$. These eigenvalues and eigenvectors extend to a complex neighborhood of any point $e^{i\xi}$ of \mathbb{S}^1 so the eigenvalues of $\mathcal{A}(\kappa)$ are geometrically regular. If $\lambda a < 1$, the leap-frog scheme enters the framework of Theorem 3.

We now turn to a more involved class of schemes:

The Runge-Kutta schemes We follow the description of Runge-Kutta schemes in [8, chapter 6]. Introducing the notation:

$$Q U_j := \frac{2}{3} (U_{j+1} - U_{j-1}) - \frac{1}{12} (U_{j+2} - U_{j-2}),$$

which is a fourth-order approximation of the space derivative $\partial_x U$, the Runge-Kutta scheme of order 3 reads:

$$u_j^{n+1} = \sum_{\ell=0}^3 \frac{(-\lambda a Q)^\ell}{\ell!} u_j^n. \quad (35)$$

We compute:

$$\mathcal{A}(\kappa) = \sum_{\ell=0}^3 \frac{(-\lambda a \widehat{Q}(\kappa))^\ell}{\ell!}, \quad \widehat{Q}(\kappa) = \frac{2}{3} (\kappa - \kappa^{-1}) - \frac{1}{12} (\kappa^2 - \kappa^{-2}).$$

Assumption 1 is satisfied as long as $a \neq 0$. We thus check the ℓ^2 -stability of the scheme and compute:

$$|\mathcal{A}(e^{i\xi})|^2 = 1 - \frac{\lambda^4 a^4}{972} h(\xi)^4 \left(1 - \frac{\lambda^2 a^2}{27} h(\xi)^2 \right), \quad h(\xi) := (4 - \cos \xi) \sin \xi.$$

The maximum of $|h|$ on \mathbb{R} can be explicitly computed and equals $|h(\xi_0)|$ where ξ_0 satisfies $\cos \xi_0 = 1 - \sqrt{6}/2$. For later use, we let β denote the maximum¹⁰ of $|h|$ on \mathbb{R} . Then the scheme (35) is ℓ^2 -stable (and geometrically regular) if and only if $\lambda a \leq 3\sqrt{3}/\beta$. We now analyze the behavior of $\mathcal{A}(\kappa)$ when it touches the unit circle \mathbb{S}^1 .

If $\lambda a < 3\sqrt{3}/\beta$, we have $\mathcal{A}(\kappa) \in \mathbb{S}^1$ if and only if $\kappa = \pm 1$. In that case, we have $\mathcal{A}'(1) = -\lambda a \neq 0$ and $\mathcal{A}'(-1) = -(5/3)\lambda a \neq 0$. If $\lambda a = 3\sqrt{3}/\beta$, we have $\mathcal{A}(\kappa) \in \mathbb{S}^1$ if and only if

¹⁰The exact value of β is $(3 + \sqrt{6}/2) \sqrt{\sqrt{6} - 3/2}$.

$\kappa = \pm 1$ or $\kappa = e^{\pm i\xi_0}$ with ξ_0 defined above. We still have $\mathcal{A}'(1) = -\lambda a \neq 0$ and $\mathcal{A}'(-1) = -(5/3)\lambda a \neq 0$ but now we also have $\mathcal{A}'(e^{\pm i\xi_0}) = 0$. However, we can also compute:

$$\operatorname{Re} \left(e^{\pm 2i\xi_0} \overline{\mathcal{A}(e^{\pm i\xi_0})} \mathcal{A}''(e^{\pm i\xi_0}) \right) = \frac{27}{4\beta^2} \left(3 - \frac{\sqrt{6}}{3} \right) > 0.$$

If $\lambda a \leq 3\sqrt{3}/\beta$, the Runge-Kutta scheme (35) enters the framework of Theorem 3.

The Runge-Kutta scheme of order 4 reads:

$$u_j^{n+1} = \sum_{\ell=0}^4 \frac{(-\lambda a Q)^\ell}{\ell!} u_j^n, \quad (36)$$

and we compute:

$$\mathcal{A}(\kappa) = \sum_{\ell=0}^4 \frac{(-\lambda a \widehat{Q}(\kappa))^\ell}{\ell!}.$$

Assumption 1 is again satisfied as long as $a \neq 0$. We thus check the ℓ^2 -stability of the scheme and compute:

$$|\mathcal{A}(e^{i\xi})|^2 = 1 - \frac{\lambda^6 a^6}{52488} h(\xi)^6 \left(1 - \frac{\lambda^2 a^2}{72} h(\xi)^2 \right),$$

where the function h is defined as above. Then the scheme (36) is ℓ^2 -stable (and geometrically regular) if and only if $\lambda a \leq 6\sqrt{2}/\beta$. We now analyze the behavior of $\mathcal{A}(\kappa)$ when it touches the unit circle.

If $\lambda a < 6\sqrt{2}/\beta$, we have $\mathcal{A}(\kappa) \in \mathbb{S}^1$ if and only if $\kappa = \pm 1$. In that case, we have $\mathcal{A}'(1) = -\lambda a \neq 0$ and $\mathcal{A}'(-1) = -(5/3)\lambda a \neq 0$. If $\lambda a = 6\sqrt{2}/\beta$, we have $\mathcal{A}(\kappa) \in \mathbb{S}^1$ if and only if $\kappa = \pm 1$ or $\kappa = e^{\pm i\xi_0}$ with ξ_0 defined above. We still have $\mathcal{A}'(1) = -\lambda a \neq 0$ and $\mathcal{A}'(-1) = -(5/3)\lambda a \neq 0$ but now we also have $\mathcal{A}'(e^{\pm i\xi_0}) = 0$. However, we can also compute:

$$\operatorname{Re} \left(e^{\pm 2i\xi_0} \overline{\mathcal{A}(e^{\pm i\xi_0})} \mathcal{A}''(e^{\pm i\xi_0}) \right) = \frac{64}{\beta^2} \left(3 - \frac{\sqrt{6}}{3} \right) > 0.$$

If $\lambda a \leq 6\sqrt{2}/\beta$, the Runge-Kutta scheme (36) enters the framework of Theorem 3. We have thus obtained two new situations where our analysis applies and that could not be analyzed with the theory in [9]. To illustrate the computations above, we show in figure 1 the curve $\{\mathcal{A}(\kappa), \kappa \in \mathbb{S}^1\}$. It touches the unit circle at a regular point and at two cusps that correspond to $\kappa = e^{\pm 2i\xi_0}$. For ξ close to ξ_0 , we have an estimate of the form:

$$|\mathcal{A}(e^{i\xi})| \leq 1 - c(\xi - \xi_0)^2.$$

A An elementary proof of Lemma 2

We assume that the matrix $\mathcal{A}(\kappa)$ defined by (4) is uniformly power bounded for $\kappa \in \mathbb{S}^1$, and that assumption 1 is satisfied. Applying the Kreiss matrix Theorem¹¹, there exists a constant $C > 0$ such that:

$$\forall \kappa \in \mathbb{S}^1, \quad \forall z \in \mathcal{U}, \quad \|(\mathcal{A}(\kappa) - zI)^{-1}\| \leq \frac{C}{|z| - 1}.$$

¹¹Here we only use the easy part of the Kreiss matrix Theorem so our proof is really elementary !

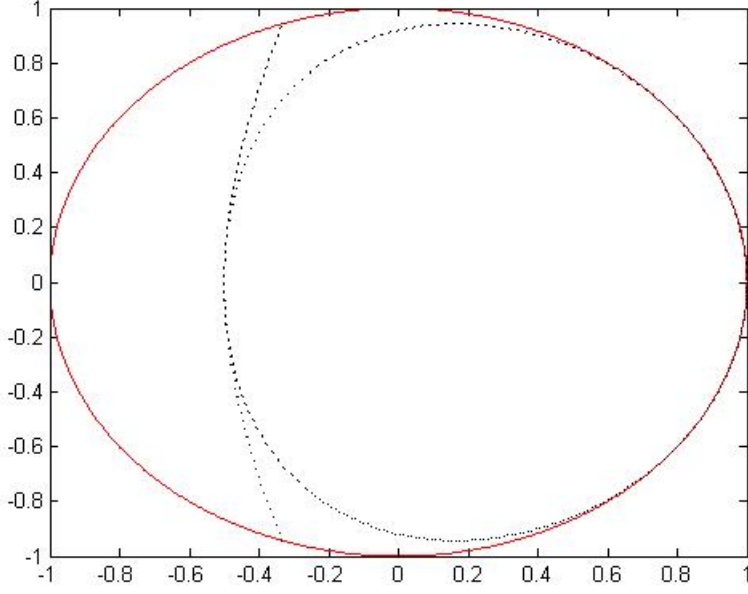


Figure 1: The curve $\{\mathcal{A}(\kappa), \kappa \in \mathbb{S}^1\}$ (black dots), and the unit circle (red).

Using the expression (4) of $\mathcal{A}(\kappa)$, we can compute the vector $(\mathcal{A}(\kappa) - zI)^{-1}Y$, with $Y = (y, 0, \dots, 0) \in \mathbb{C}^{N(s+1)}$ and $y \in \mathbb{C}^N$. Indeed, let $X = (x_0, \dots, x_s) \in \mathbb{C}^{N(s+1)}$ be the unique solution to $(\mathcal{A}(\kappa) - zI)X = Y$. We have:

$$\begin{aligned} \forall \sigma = 0, \dots, s, \quad x_\sigma &= z^{s-\sigma} x_s, \\ \left(I - \sum_{\sigma=0}^s z^{-\sigma-1} \widehat{Q}_\sigma(\kappa) \right) x_s &= -z^{-s-1} y. \end{aligned}$$

We have the estimate $(|z| - 1)|X| \leq C|y|$ so in particular, we have $(|z| - 1)|x_0| \leq C|y|$. Using the relation $x_0 = z^s x_s$, we get an estimate:

$$|x_s| \leq \frac{C|z|^{-s}}{|z| - 1} |y|, \quad \text{where } x_s = -z^{-s-1} \left(I - \sum_{\sigma=0}^s z^{-\sigma-1} \widehat{Q}_\sigma(\kappa) \right)^{-1} y.$$

Taking the supremum over $y \in \mathbb{C}^N$, we obtain that there exists a constant $C > 0$ such that:

$$\forall \kappa \in \mathbb{S}^1, \quad \forall z \in \mathcal{U}, \quad \left\| \left(I - \sum_{\sigma=0}^s z^{-\sigma-1} \widehat{Q}_\sigma(\kappa) \right)^{-1} \right\| \leq C \frac{|z|}{|z| - 1}. \quad (37)$$

We also easily compute the relation:

$$I - \sum_{\sigma=0}^s z^{-\sigma-1} \widehat{Q}_\sigma(\kappa) = \sum_{\ell=-r}^p \kappa^\ell \mathbb{A}_\ell(z), \quad (38)$$

where the matrices $\mathbb{A}_\ell(z)$ are defined by (7).

Let $\kappa \in \mathbb{S}^1$ and let $z \in \mathcal{U}$. We now consider a vector $b = (b_p, \dots, b_{-r+1}) \in \mathbb{C}^{N(p+r)}$ and we let $x = (x_{p-1}, \dots, x_{-r})$ be the unique solution to $(\mathbb{M}(z) - \kappa I)x = b$. From the definition (8), we obtain the relations:

$$\forall \ell = -r+1, \dots, p-1, \quad x_\ell = \kappa^{r+\ell} x_{-r} + \sum_{j=0}^{\ell+r-1} \kappa^j b_{\ell-j},$$

$$\kappa^r \left(\sum_{\ell=-r}^p \kappa^\ell \mathbb{A}_\ell(z) \right) x_{-r} = -\tilde{b}(\kappa, z),$$

with a vector $\tilde{b}(\kappa, z)$ defined by:

$$\tilde{b}(\kappa, z) := \mathbb{A}_p(z) b_p + \sum_{\ell=-r+1}^{p-1} \mathbb{A}_\ell(z) \sum_{j=0}^{\ell+r-1} \kappa^j b_{\ell-j} + \kappa \mathbb{A}_p(z) \sum_{j=0}^{p+r-2} \kappa^j b_{p-1-j}.$$

For $z \in \mathcal{U}$ and $\kappa \in \mathbb{S}^1$, we have a uniform bound:

$$|\tilde{b}(\kappa, z)| \leq C |b|,$$

because the matrices $\mathbb{A}_\ell(z)$ are uniformly bounded for $z \in \mathcal{U}$, see (7). We then use the relation (38) and the estimate (37) to obtain an upper bound for $|x_{-r}|$:

$$|x_{-r}| \leq C \frac{|z|}{|z|-1} |b|,$$

with a constant C that is uniform with respect to $\kappa \in \mathbb{S}^1$ and $z \in \mathcal{U}$. The other components x_{-r+1}, \dots, x_{p-1} of x are easily estimated in terms of x_{-r} and b . We have thus proved that there exists a constant $C > 0$ such that for all $z \in \mathcal{U}$ and for all $\kappa \in \mathbb{S}^1$, we have:

$$|(\mathbb{M}(z) - \kappa I)^{-1} b| \leq C \frac{|z|}{|z|-1} |b|.$$

The proof of Lemma 2 is complete.

B Proof of Lemma 6

We already know that $\mathcal{L}(E)$ is a Banach space and that the set of isomorphisms $Gl(E)$ is an open subset of $\mathcal{L}(E)$. This first property shows that the set $\{t \in \mathcal{T} / L(t) \in Gl(E)\}$ is open because L is continuous. It only remains to show that this set is closed and the claim will follow (this set is nonempty thanks to the assumption of Lemma 6). We thus consider a sequence (t_n) in \mathcal{T} that converges to some $\underline{t} \in \mathcal{T}$ and such that for all n , $L(t_n)$ belongs to $Gl(E)$. We are going to show that $L(\underline{t})$ also belongs to $Gl(E)$. Using the Banach isomorphism Theorem, it is enough to prove that $L(\underline{t})$ is a bijection.

Due to the uniform bound $|x|_E \leq C_0 |L(t)x|_E$, it is clear that $L(\underline{t})$ is injective and that for all n we have $\|L(t_n)^{-1}\|_{\mathcal{L}(E)} \leq C_0$. It remains to show that $L(\underline{t})$ is surjective. Let $y \in E$. For all integers n and p , we have:

$$\begin{aligned} |L(t_{n+p})^{-1} y - L(t_n)^{-1} y|_E &\leq \|L(t_{n+p})^{-1} - L(t_n)^{-1}\|_{\mathcal{L}(E)} |y|_E \\ &\leq \|L(t_{n+p})^{-1} (L(t_n) - L(t_{n+p})) L(t_n)^{-1}\|_{\mathcal{L}(E)} |y|_E \leq C_0^2 \|L(t_{n+p}) - L(t_n)\|_{\mathcal{L}(E)} |y|_E. \end{aligned}$$

These inequalities show that $(L(t_n)^{-1} y)$ is a Cauchy sequence in E and therefore converges to some $x \in E$. Moreover we have $L(t_n) L(t_n)^{-1} y = y$ for all n and passing to the limit we get $L(\underline{t}) x = y$. Here we use again the continuity of L . This shows that $L(\underline{t})$ is surjective, which completes the proof.

C An extension of Theorem 3

The aim of this appendix is to prove an extended version of Theorem 3 that allows higher dissipation estimates than the second order dissipation that corresponds to case *iii*) in Theorem 3.

C.1 A preliminary fact on parametrized curves

We first need to establish the following:

Lemma 9. *Let $\underline{\kappa} \in \mathbb{S}^1$, let \mathcal{W} be an open neighborhood of $\underline{\kappa}$, and let λ be a holomorphic function defined on \mathcal{W} with $\lambda(\underline{\kappa}) \in \mathbb{S}^1$. Assume moreover that for all $\xi \in \mathbb{R}$ sufficiently close to 0, we have $|\lambda(\underline{\kappa} e^{i\xi})| \leq 1$. Then one of the two following properties holds true:*

- *for all $\xi \in \mathbb{R}$ sufficiently close to 0, we have $|\lambda(\underline{\kappa} e^{i\xi})| = 1$,*
- *there exists a positive integer k and a constant $c > 0$ such that for all $\xi \in \mathbb{R}$ sufficiently close to 0, we have:*

$$|\lambda(\underline{\kappa} e^{i\xi})| \leq 1 - c\xi^{2k}. \quad (39)$$

Lemma 9 shows that the parametrized curve $\{\lambda(\underline{\kappa} e^{i\xi}), \xi \in \mathbb{R}\}$ can be tangent to infinite order to the unit circle if and only if it is included in the unit circle. In all other cases, the tangency point corresponds to a regular point ($\lambda'(\underline{\kappa}) \neq 0$) or to a cusp ($\lambda'(\underline{\kappa}) = 0$). We refer to figure 1 for an example.

Proof of Lemma 9. The function:

$$f : \zeta \longmapsto \lambda(\underline{\kappa} e^{i\zeta}),$$

is holomorphic on some neighborhood of 0, and $f(0) \in \mathbb{S}^1$. We can therefore define a holomorphic function g on some neighborhood of 0 such that $f = e^g$. We then have $\operatorname{Re} g = \ln |f|$. Eventually, let us define the real valued \mathcal{C}^∞ function $h(\xi) = |f(\xi)|$ for $\xi \in \mathbb{R}$ sufficiently close to 0.

Let us assume that all derivatives of h at 0 vanish (here the derivatives are taken with respect to the real variable ξ). Then all derivatives of $\ln h$ with respect to ξ at 0 vanish:

$$\forall k \in \mathbb{N}, \quad \frac{d^k \ln h}{d\xi^k}(0) = \frac{\partial^k \operatorname{Re} g}{\partial \xi^k}(0) = \operatorname{Re} \frac{\partial^k g}{\partial \xi^k}(0) = 0.$$

The Cauchy-Riemann relations imply that all derivatives of g with respect to the complex variable ζ at 0 are purely imaginary, so $g(\xi) \in i\mathbb{R}$ for all $\xi \in \mathbb{R}$ close to 0 (use an expansion of g in power series). We have therefore obtained $f(\xi) \in \mathbb{S}^1$ for all $\xi \in \mathbb{R}$ close to 0.

The only other possibility for h is that there exists a smallest positive integer k_0 such that $h^{(k_0)}(0) \neq 0$. Thanks to the assumption on λ , we have $h(\xi) \leq 1$ for all ξ close to 0. This property implies that k_0 is even and $h^{(k_0)}(0) < 0$. The claim follows from Taylor's formula. \square

Using Taylor's formula, we see that the case $k = 1$ in (39) corresponds either to condition *i*) or to condition *iii*) in Theorem 3. In the following paragraph, we show how to take the case $k = 2$ into account in the symmetrizers construction.

C.2 A generalization of Theorem 3

We now prove the following:

Theorem 4. *Let assumption 1 be satisfied, and assume that the symbol $\mathcal{A}(\kappa)$ is uniformly power bounded for $\kappa \in \mathbb{S}^1$. Assume moreover that all the eigenvalues of $\mathcal{A}(\kappa)$, $\kappa \in \mathbb{S}^1$, that belong to \mathbb{S}^1 are geometrically regular and that at least one of the five following properties is satisfied by each eigenvalue $\lambda_j(\kappa)$ in the decomposition (10):*

- i) $\lambda_j'(\underline{\kappa}) \neq 0$,
- ii) $\lambda_j(\kappa) \in \mathbb{S}^1$ for all $\kappa \in \mathbb{S}^1$ sufficiently close to $\underline{\kappa}$,
- iii) $\operatorname{Re}(\underline{\kappa}^2 \overline{\lambda_j(\underline{\kappa})} \lambda_j''(\underline{\kappa})) > 0$.
- iv) $\underline{\kappa}^2 \overline{\lambda_j(\underline{\kappa})} \lambda_j''(\underline{\kappa}) \in i\mathbb{R} \setminus \{0\}$ and $\operatorname{Re} \left[\overline{\lambda_j(\underline{\kappa})} (\underline{\kappa}^4 \lambda_j^{(4)}(\underline{\kappa}) + 6 \underline{\kappa}^3 \lambda_j'''(\underline{\kappa})) \right] + 3 |\lambda_j''(\underline{\kappa})|^2 < 0$.
- v) $\lambda_j'(\underline{\kappa}) = \lambda_j''(\underline{\kappa}) = \lambda_j'''(\underline{\kappa}) = 0$ and $\operatorname{Re}(\underline{\kappa}^4 \overline{\lambda_j(\underline{\kappa})} \lambda_j^{(4)}(\underline{\kappa})) < 0$.

Then \mathbb{M} defined by (8) admits a K -symmetrizer. In particular, if the uniform Kreiss-Lopatinskii condition is satisfied, the scheme (2) is strongly stable.

Proof. We follow the proof of Theorem 3. In the block decomposition of $\mathbb{M}(z)$, we can have some new 2×2 blocks that are associated with an eigenvalue λ_j of \mathcal{A} satisfying condition iv) in Theorem 4 or some 4×4 blocks that are associated with an eigenvalue λ_j satisfying condition v). We thus only need to show that we can construct a K -symmetrizer for such blocks.

Let us first deal with a 2×2 matrix $P(z)$ that satisfies:

$$\det(P(z) - \kappa I) = \vartheta(\kappa, z) (z - \lambda(\kappa)), \quad P(\underline{z}) = \underline{\kappa} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix},$$

where λ is holomorphic in a neighborhood of $\underline{\kappa}$ and satisfies condition iv) in Theorem 4 (we drop the index j for simplicity). Let us begin with the following:

Lemma 10. *Let $\underline{\kappa} = e^{i\xi}$ and let us define the holomorphic function $\ell(\zeta) := \ln [\lambda(e^\zeta)/\underline{z}]$ for ζ close to $i\xi$. Then we have:*

$$\ell'(i\xi) = 0, \quad \ell''(i\xi) \in i\mathbb{R} \setminus \{0\}, \quad \ell'''(i\xi) \in \mathbb{R}, \quad \operatorname{Re} \ell^{(4)}(i\xi) < 0.$$

The proof of Lemma 10 follows from the assumptions on λ and from straightforward computations so we omit it. As in the proof of Theorem 3, we can then define a matrix $\mathbb{P}(\tau)$ that depends holomorphically on τ in a neighborhood of 0 and such that $P(z) = \exp \mathbb{P}(\ln(z/\underline{z}))$. The lower left coefficient of $\mathbb{P}'(0)$ is nonzero thanks to (30) so we have:

$$\det(\mathbb{P}(\tau) - \zeta I) = \vartheta(\zeta, \tau) (\tau - \ell(\zeta)),$$

where ℓ is defined in Lemma 10. Moreover, there exists a holomorphic change of basis $Q(\tau)$ such that:

$$\check{\mathbb{P}}(\tau) := Q(\tau)^{-1} \mathbb{P}(\tau) Q(\tau) = \begin{pmatrix} i\xi & i \\ 0 & i\xi \end{pmatrix} + \begin{pmatrix} b_1(\tau) & 0 \\ b_2(\tau) & 0 \end{pmatrix}, \quad (40)$$

where both functions b_1, b_2 vanish for $\tau = 0$. Using Lemma 10 and differentiating four times the relation:

$$(\zeta - i\xi)^2 = (\zeta - i\xi) b_1(\ell(\zeta)) + i b_2(\ell(\zeta)),$$

we obtain the following result:

$$b'_2(0) \in \mathbb{R} \setminus \{0\}, \quad b'_1(0) \in \mathbb{R}, \quad b'_2(0) \operatorname{Re} b''_2(0) < 0. \quad (41)$$

Then the proof of Theorem 4 follows from:

Lemma 11. *Let $\check{\mathbb{P}}$ denote the matrix in (40) and let the functions b_1, b_2 satisfy (41). Let $K \geq 1$. Then there exist a neighborhood of 0, a \mathcal{C}^∞ mapping \check{S} defined on this neighborhood with values in $\mathcal{M}_2(\mathbb{C})$, and there exists a constant $c > 0$ such that for all τ close to 0 the following properties hold:*

- $\check{S}(\tau)$ is hermitian,
- $\operatorname{Re} (\check{S}(\tau) \check{\mathbb{P}}(\tau)) \geq c (\operatorname{Re} \tau + (\operatorname{Im} \tau)^2) I$ if $\operatorname{Re} \tau \geq 0$,
- for all $W \in \mathbb{C}^2$, $W^* \check{S}(0) W \geq (K^2 + 1/2) |W_2|^2 - |W_1|^2/2$.

Using Lemma 11, the proof of Theorem 4 follows with the same arguments as used in the proof of Theorem 3. We thus focus on the proof of Lemma 11.

Proof of Lemma 11. We introduce the notations:

$$\alpha_1 := b'_1(0), \quad \alpha_2 := b'_2(0), \quad \beta := -\frac{b''_2(0)}{\alpha_2}.$$

From (41) the number β has positive real part. As in the proof of Lemma 8, we use the notation $\tau = \gamma + i\delta$. We can write:

$$b_1(i\delta) = i\alpha_1\delta + \delta^2 g_1(\delta), \quad b_2(i\delta) = i\alpha_2\delta + \alpha_2\beta\delta^2 + \delta^3 g_2(\delta),$$

for some appropriate analytic functions g_1, g_2 . The symmetrizer \check{S} is chosen under the form:

$$\check{S}(\tau) = \check{S}_0(\delta) + \gamma \check{S}_1,$$

where \check{S}_0 is a \mathcal{C}^∞ function of the real variable δ , and \check{S}_1 is a constant hermitian matrix. The matrix $\check{S}_0(\delta)$ is chosen under the form¹²:

$$\check{S}_0(\delta) := \begin{pmatrix} \alpha_2(\alpha_1 + b)\delta & \alpha_2 + ie\delta^2 \\ \alpha_2 - ie\delta^2 & b \end{pmatrix},$$

where b, e are real constants. We first fix $b > 0$ sufficiently large so that the following estimate holds:

$$W^* \check{S}_0(0) W \geq (K^2 + \frac{1}{2}) |W_2|^2 - \frac{1}{2} |W_1|^2.$$

With our choice for \check{S}_0 , we compute:

$$\operatorname{Re} (\check{S}_0(\delta) \check{\mathbb{P}}(i\delta)) = \delta^2 \begin{pmatrix} \alpha_2^2 \operatorname{Re} \beta & \star \\ \frac{\alpha_2}{2} (b\beta + g_1(0)) & e \end{pmatrix} + O(\delta^3),$$

¹²The expression for \check{S}_0 differs from the one used in the proof of Lemma 8, and it also differs from the construction used by Kreiss in the continuous case. For each dissipative behavior of the eigenvalue λ_j there corresponds a specific form of the symmetrizer.

where the \star coefficient is such that the matrix above is hermitian. We can then choose e large enough such that for all δ sufficiently small, we have:

$$\operatorname{Re} (\check{S}_0(\delta) \check{\mathbb{P}}(i\delta)) \geq c\delta^2 I,$$

for some appropriate constant $c > 0$. The construction of \check{S} is then achieved, as in Lemma 8, by taking:

$$\check{S}_1 := \begin{pmatrix} 0 & ig \\ -ig & 0 \end{pmatrix},$$

with $g > 0$ sufficiently large. The proof of Lemma 11 is complete. \square

Up to now, we have proved the existence of a K -symmetrizer for blocks associated with eigenvalues λ_j that satisfy condition *iv*) in Theorem 4. To complete the proof of Theorem 4, we turn to the case of eigenvalues λ_j that satisfy condition *v*). More precisely, we consider a 4×4 matrix $P(z)$ that satisfies:

$$\det(P(z) - \kappa I) = \vartheta(\kappa, z)(z - \lambda(\kappa)), \quad P(\underline{z}) = \underline{\kappa} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where λ is holomorphic in a neighborhood of $\underline{\kappa}$ and satisfies condition *v*) in Theorem 4 (we drop again the index j for simplicity). As in Lemma 10, we introduce the holomorphic function $\ell(\zeta) := \ln[\lambda(e^\zeta)/\underline{z}]$ for ζ close to $i\underline{\xi}$ ($\underline{\xi}$ denotes the argument of $\underline{\kappa}$ as in Lemma 10). Then we have:

$$\ell'(i\underline{\xi}) = \ell''(i\underline{\xi}) = \ell'''(i\underline{\xi}) = 0, \quad \operatorname{Re} \ell^{(4)}(i\underline{\xi}) < 0. \quad (42)$$

As in the proof of Theorem 3, we can define a matrix $\mathbb{P}(\tau)$ that depends holomorphically on τ in a neighborhood of 0 and such that $P(z) = \exp \mathbb{P}(\ln(z/\underline{z}))$. The lower left coefficient of $\mathbb{P}'(0)$ is nonzero (use again (30)) and we have:

$$\det(\mathbb{P}(\tau) - \zeta I) = \vartheta(\zeta, \tau)(\tau - \ell(\zeta)), \quad (43)$$

where ℓ is defined above and satisfies (42). Moreover, Ralston's Lemma [19] shows that there exists a holomorphic change of basis $Q(\tau)$ such that:

$$\check{\mathbb{P}}(\tau) := Q(\tau)^{-1} \mathbb{P}(\tau) Q(\tau) = \begin{pmatrix} i\underline{\xi} & i & 0 & 0 \\ 0 & i\underline{\xi} & i & 0 \\ 0 & 0 & i\underline{\xi} & i \\ 0 & 0 & 0 & i\underline{\xi} \end{pmatrix} + \begin{pmatrix} b_1(\tau) & 0 & 0 & 0 \\ b_2(\tau) & 0 & 0 & 0 \\ b_3(\tau) & 0 & 0 & 0 \\ b_4(\tau) & 0 & 0 & 0 \end{pmatrix}, \quad (44)$$

where all functions b_1, b_2, b_3, b_4 vanish for $\tau = 0$. Using (43), we have $\det(\mathbb{P}(\ell(\zeta)) - \zeta I) = 0$ for all ζ . Expanding this 4×4 determinant and using the property (42) for ℓ , we find:

$$\operatorname{Im} b'_4(0) < 0. \quad (45)$$

The existence of a K -symmetrizer is summarized in the following:

Lemma 12. *Let $\check{\mathbb{P}}$ denote the matrix in (44) and let the functions b_1, b_2, b_3, b_4 satisfy (45). Let $K \geq 1$. Then there exist a neighborhood of 0, a \mathcal{C}^∞ mapping \check{S} defined on this neighborhood with values in $\mathcal{M}_4(\mathbb{C})$, and there exists a constant $c > 0$ such that for all τ close to 0 the following properties hold:*

- $\check{S}(\tau)$ is hermitian,
- $\operatorname{Re} (\check{S}(\tau) \check{\mathbb{P}}(\tau)) \geq c (\operatorname{Re} \tau + (\operatorname{Im} \tau)^2) I$ if $\operatorname{Re} \tau \geq 0$,
- for all $W \in \mathbb{C}^4$, $W^* \check{S}(0) W \geq (K^2 + 1/2) |W_2|^2 - |W_1|^2/2$ where we use the decomposition $W = (W_1, W_2)$, $W_1, W_2 \in \mathbb{C}^2$.

Proof of Lemma 12. We keep the notation $\tau = \gamma + i\delta$. In the proof of Lemma 12, we shall frequently decompose 4×4 matrices into four subblocks of size 2×2 . For instance, the matrix $\check{\mathbb{P}}(i\delta)$ is decomposed as follows:

$$\check{\mathbb{P}}(i\delta) = i\xi I + \begin{pmatrix} \mathbb{P}_1 & -\mathbb{P}_1^* \\ 0 & \mathbb{P}_1 \end{pmatrix} + \begin{pmatrix} \delta B_1(\delta) & 0 \\ i\delta B_2 + \delta^2 B_3(\delta) & 0 \end{pmatrix}, \quad \mathbb{P}_1 := \begin{pmatrix} 0 & i \\ 0 & 0 \end{pmatrix}, \quad (46)$$

where B_1, B_3 are analytic functions with values in $\mathcal{M}_2(\mathbb{C})$, and B_2 is the constant matrix defined by:

$$B_2 := \begin{pmatrix} b'_3(0) & 0 \\ b'_4(0) & 0 \end{pmatrix}. \quad (47)$$

We seek the symmetrizer \check{S} under the form:

$$\check{S}(\tau) = \check{S}_0(\delta) + \gamma \check{S}_1,$$

where \check{S}_0 is a \mathcal{C}^∞ function of the real variable δ , and \check{S}_1 is a constant hermitian matrix. Let us first focus on the construction of \check{S}_0 . We choose \check{S}_0 of the following form:

$$\check{S}_0(\delta) := \begin{pmatrix} \delta^2 A_0 & B_0 + \delta C_0 \\ B_0^* + \delta C_0^* & D_0 \end{pmatrix}, \quad (48)$$

where A_0, D_0 are constant hermitian 2×2 matrices, and where B_0, C_0 are constant 2×2 matrices to be chosen later on. Using (46) and (48), we obtain:

$$\begin{aligned} \check{S}_0(\delta) \check{\mathbb{P}}(i\delta) = & \begin{pmatrix} i\delta B_0 B_2 + \delta^2 (A_0 \mathbb{P}_1 + i C_0 B_2 + B_0 B_3(0)) & B_0 \mathbb{P}_1 + \delta C_0 \mathbb{P}_1 \\ B_0^* \mathbb{P}_1 + \delta (C_0^* \mathbb{P}_1 + B_0^* B_1(0) + i D_0 B_2) & D_0 \mathbb{P}_1 - B_0^* \mathbb{P}_1^* \end{pmatrix} \\ & + \begin{pmatrix} O(\delta^3) & O(\delta^2) \\ O(\delta^2) & O(\delta) \end{pmatrix}, \end{aligned} \quad (49)$$

and then we take the real part of this equality. First of all, we choose the matrix B_0 such that when we compute the real part, the extra-diagonal term vanishes when $\delta = 0$. This is equivalent to requiring $B_0^* \mathbb{P}_1 + (B_0 \mathbb{P}_1)^* = 0$, so we choose B_0 of the form:

$$B_0 := \begin{pmatrix} 0 & e_1 \\ e_1 & e_2 \end{pmatrix}, \quad e_1, e_2 \in \mathbb{C}. \quad (50)$$

Next, we want to make the linear term in δ in the upper left subblock of (49) vanish. In other words, we want $B_0 B_2$ to be hermitian. Recalling that B_2 is given by (47), and B_0 is given by (50), we choose:

$$e_1 := \alpha \overline{b'_4(0)}, \quad \alpha \in \mathbb{R}, \quad e_2 := -\frac{e_1 b'_3(0)}{b'_4(0)}. \quad (51)$$

Our next requirement is that the lower right subblock in the real part of (49) should be definite positive, uniformly with respect to δ . This amounts to choosing B_0 and D_0 such that the real part of $D_0 \mathbb{P}_1 - B_0^* \mathbb{P}_1^*$ is definite positive. We compute:

$$\operatorname{Re} (D_0 \mathbb{P}_1 - B_0^* \mathbb{P}_1^*) = \begin{pmatrix} \operatorname{Im} e_1 & i(a - e_2)/2 \\ i(\overline{e_2} - a)/2 & \operatorname{Im} e_3 \end{pmatrix}, \quad \text{with } D_0 := \begin{pmatrix} a & e_3 \\ \overline{e_3} & b \end{pmatrix}. \quad (52)$$

We recall that the coefficient $b'_4(0)$ satisfies (45), so we choose $\alpha = 1$ in (51). The matrix B_0 is then completely determined by (50) and (51). Part of the matrix D_0 will be determined by the “boundary conditions” and part of it will be determined by (52). More precisely, let us first fix the coefficient $a = K^2 + 1 + \|B_0\|^2$ in the definition (52) of D_0 . Then we can choose $e_3 \in i\mathbb{R}$ such that the matrix $\text{Re}(D_0 \mathbb{P}_1 - B_0^* \mathbb{P}_1^*)$ in (52) is positive definite. Eventually, we can choose the lower right coefficient b of D_0 such that the matrix $\check{S}(0)$ satisfies the last condition of Lemma 12, see (48). (Recall that our definition of \check{S} gives $\check{S}(0) = \check{S}_0(0)$ and $\check{S}_0(0)$ is given in (48), with B_0 given by (50), (51).)

Let us summarize what we have done above. We have determined the matrices B_0 and D_0 such that the last condition in Lemma 12 is satisfied, and such that taking the real part of (49), we obtain:

$$\text{Re}(\check{S}_0(\delta) \check{\mathbb{P}}(i\delta)) = \begin{pmatrix} \delta^2 \text{Re}(A_0 \mathbb{P}_1 + i C_0 B_2 + B_0 B_3(0)) & \star \\ \delta(C_0^* \mathbb{P}_1 + (C_0 \mathbb{P}_1)^* + B_0^* B_1(0) + i D_0 B_2)/2 & \text{Re}(D_0 \mathbb{P}_1 - B_0^* \mathbb{P}_1^*) \end{pmatrix} + \begin{pmatrix} O(\delta^3) & \star \\ O(\delta^2) & O(\delta) \end{pmatrix}. \quad (53)$$

The lower right 2×2 matrix in (53) is positive definite, uniformly with respect to δ provided that δ is sufficiently small. We now choose C_0 such that $C_0^* \mathbb{P}_1 + (C_0 \mathbb{P}_1)^*$ vanishes, so the lower left 2×2 matrix will be completely determined by B_0 and D_0 that have been fixed above. It is sufficient to choose C_0 of the form:

$$C_0 := \begin{pmatrix} 0 & e_4 \\ e_4 & e_5 \end{pmatrix}, \quad e_4, e_5 \in \mathbb{C}, \quad (54)$$

so (53) reduces to:

$$\text{Re}(\check{S}_0(\delta) \check{\mathbb{P}}(i\delta)) = \begin{pmatrix} \delta^2 \text{Re}(A_0 \mathbb{P}_1 + i C_0 B_2 + B_0 B_3(0)) & \star \\ \delta(B_0^* B_1(0) + i D_0 B_2)/2 & \text{Re}(D_0 \mathbb{P}_1 - B_0^* \mathbb{P}_1^*) \end{pmatrix} + \begin{pmatrix} O(\delta^3) & \star \\ O(\delta^2) & O(\delta) \end{pmatrix}. \quad (55)$$

Choosing C_0 as in (54), and A_0 of the form¹³:

$$A_0 := \begin{pmatrix} 0 & \overline{e_6} \\ e_6 & 0 \end{pmatrix},$$

we compute:

$$\text{Re}(A_0 \mathbb{P}_1 + i C_0 B_2 + B_0 B_3(0)) = \underbrace{\begin{pmatrix} -\text{Im}(e_4 b'_4(0)) & \star \\ i(e_4 b'_3(0) + e_5 b'_4(0))/2 & -\text{Im} e_6 \end{pmatrix}}_H + \text{Re}(B_0 B_3(0)).$$

At this stage, we can choose e_5 such that the extra-diagonal term in H vanishes, and we choose $e_4 = \mu \in \mathbb{R}$, $e_6 = -i\mu$ with $\mu > 0$ so large that the diagonal blocks in (55) dominate the extra-diagonal blocks (recall that the extra-diagonal terms do not depend on μ). More precisely, we can achieve the estimate:

$$\text{Re}(\check{S}_0(\delta) \check{\mathbb{P}}(i\delta)) \geq \begin{pmatrix} c\delta^2 I & 0 \\ 0 & cI \end{pmatrix} + \begin{pmatrix} O(\delta^3) & \star \\ O(\delta^2) & O(\delta) \end{pmatrix}, \quad (56)$$

¹³Recall that A_0 should be hermitian because $\check{S}_0(\delta)$ should be hermitian for all δ .

where c is a positive constant, and the “remainder terms” in the right-hand side of (56) can be absorbed by choosing δ sufficiently small. The matrix $\check{S}_0(\delta)$ is now completely determined, and it satisfies:

$$\operatorname{Re} (\check{S}_0(\delta) \check{\mathbb{P}}(i\delta)) \geq c\delta^2 I.$$

The last task is to choose \check{S}_1 . We start from the relation:

$$\operatorname{Re} (\check{S}(\tau) \check{\mathbb{P}}(\tau)) = \operatorname{Re} (\check{S}_0(\delta) \check{\mathbb{P}}(i\delta)) + \gamma \operatorname{Re} (\check{S}_0(0) \check{\mathbb{P}}'(0) + \check{S}_1 \check{\mathbb{P}}(0)) + \gamma o(1). \quad (57)$$

Then we choose a hermitian matrix \check{S}_1 of the form:

$$\check{S}_1 = i \begin{pmatrix} 0 & g_1 & 0 & g_2 \\ -g_1 & 0 & g_2 & 0 \\ 0 & -g_2 & 0 & g_2 \\ -g_2 & 0 & -g_2 & 0 \end{pmatrix}, \quad g_1, g_2 \in \mathbb{R}.$$

We compute:

$$\operatorname{Re} (\check{S}_0(0) \check{\mathbb{P}}'(0) + \check{S}_1 \check{\mathbb{P}}(0)) = \begin{pmatrix} H_1 & \star \\ H_2 & g_2 I \end{pmatrix}, \quad (58)$$

where the matrix H_1 is hermitian, and H_1, H_2 are given by:

$$H_1 = \begin{pmatrix} |b'_4(0)|^2 & 0 \\ 0 & g_1 \end{pmatrix}, \quad H_2 = \frac{1}{2} (B_0^* B_1(0) + D_0 B_2) + \begin{pmatrix} -g_1/2 & 0 \\ 0 & 0 \end{pmatrix}.$$

We first choose $g_1 = 1$, then we choose $g_2 > 0$ sufficiently large in order to “absorb” the extra-diagonal block H_2 in (58). The decomposition (57) shows that the second property of Lemma 12. The proof of Lemma 12 is complete. \square

The end of the analysis follows from the same arguments used in the end of the proof of Theorem 3. More precisely, we have a K -symmetrizer for $\check{\mathbb{P}}$, then we can construct a K -symmetrizer for the 4×4 block P . The proof of Theorem 4 is complete. \square

To conclude this article, we give an example of a numerical scheme such that condition $v)$ in Theorem 4 occurs. We keep the same notations as in the section on Runge-Kutta schemes. We consider the following discretization of the space derivative:

$$QU_j := \frac{3}{8} (U_{j+1} - U_{j-1}) + \frac{1}{24} (U_{j+3} - U_{j-3}),$$

which is a second-order approximation¹⁴ of the space derivative $\partial_x U$. As in (35), the Runge-Kutta scheme of order 3 reads:

$$u_j^{n+1} = \sum_{\ell=0}^3 \frac{(-\lambda a Q)^\ell}{\ell!} u_j^n. \quad (59)$$

We compute:

$$\mathcal{A}(\kappa) = \sum_{\ell=0}^3 \frac{(-\lambda a \widehat{Q}(\kappa))^\ell}{\ell!}, \quad \widehat{Q}(\kappa) = \frac{3}{8} (\kappa - \kappa^{-1}) + \frac{1}{24} (\kappa^3 - \kappa^{-3}).$$

¹⁴It may seem absurd to consider a scheme with so many points to reach only second-order accuracy. However, if one wishes to do a general theory, the theory should be able to cover all cases, even the absurd ones !

Assumption 1 is satisfied as long as $a \neq 0$. We thus check the ℓ^2 -stability of the scheme and compute:

$$|\mathcal{A}(e^{i\xi})|^2 = 1 - \frac{\lambda^4 a^4}{12} h(\xi)^4 \left(1 - \frac{\lambda^2 a^2}{3} h(\xi)^2\right), \quad h(\xi) := \sin \xi - \frac{1}{3} \sin^3 \xi.$$

The maximum of $|h|$ on \mathbb{R} is $2/3$, and it is attained when $\xi \in \pi/2 + \mathbb{Z}\pi$. Then the scheme (59) is ℓ^2 -stable (and geometrically regular) if and only if $\lambda a \leq 3\sqrt{3}/2$.

We now analyze the behavior of $\mathcal{A}(\kappa)$ when it touches the unit circle \mathbb{S}^1 . We assume that the CFL condition is chosen in an optimal way, that is $\lambda a = 3\sqrt{3}/2$. Then we have $\mathcal{A}(\kappa) \in \mathbb{S}^1$ if and only if $\kappa \in \{\pm 1, \pm i\}$. We compute $\widehat{Q}(i) = 2i/3$, $\widehat{Q}'(i) = \widehat{Q}''(i) = \widehat{Q}'''(i) = 0$, and $\widehat{Q}^{(4)}(i) = -6i$. We then derive $\mathcal{A}'(i) = \mathcal{A}''(i) = \mathcal{A}'''(i) = 0$, and $\mathcal{A}^{(4)}(i) = 27 - i9\sqrt{3}/2$. Then we can check that condition v) in Theorem 4 is satisfied at $\kappa = i$. The same conclusion holds at $\kappa = -i$.

References

- [1] H. Baumgärtel. *Analytic perturbation theory for matrices and operators*. Birkhäuser Verlag, 1985.
- [2] J. Chazarain, A. Piriou. *Introduction to the theory of linear partial differential equations*. North-Holland, 1982.
- [3] Y. Coudière, J.-P. Vila, and P. Villedieu. Convergence d'un schéma volumes finis explicite en temps pour les systèmes hyperboliques linéaires symétriques en domaines bornés. *C. R. Acad. Sci. Paris Sér. I Math.*, 331(1):95–100, 2000.
- [4] E. Godlewski, P.-A. Raviart. *Numerical approximation of hyperbolic systems of conservation laws*. Springer-Verlag, 1996.
- [5] M. Goldberg, E. Tadmor. Scheme-independent stability criteria for difference approximations of hyperbolic initial-boundary value problems. II. *Math. Comp.*, 36(154):603–626, 1981.
- [6] M. Goldberg, E. Tadmor. Convenient stability criteria for difference approximations of hyperbolic initial-boundary value problems. II. *Math. Comp.*, 48(178):503–520, 1987.
- [7] B. Gustafsson. The convergence rate for difference approximations to mixed initial boundary value problems. *Math. Comp.*, 29(130):396–406, 1975.
- [8] B. Gustafsson, H.-O. Kreiss, and J. Olinger. *Time dependent problems and difference methods*. John Wiley & Sons, 1995.
- [9] B. Gustafsson, H.-O. Kreiss, and A. Sundström. Stability theory of difference approximations for mixed initial boundary value problems. II. *Math. Comp.*, 26(119):649–686, 1972.
- [10] H.-O. Kreiss. Stability theory for difference approximations of mixed initial boundary value problems. I. *Math. Comp.*, 22:703–714, 1968.
- [11] H.-O. Kreiss. Initial boundary value problems for hyperbolic systems. *Comm. Pure Appl. Math.*, 23:277–298, 1970.

- [12] A. Majda, S. Osher. Initial-boundary value problems for hyperbolic equations with uniformly characteristic boundary. *Comm. Pure Appl. Math.*, 28(5):607–675, 1975.
- [13] G. Métivier. The block structure condition for symmetric hyperbolic problems. *Bull. London Math. Soc.*, 32:689–702, 2000.
- [14] G. Métivier, K. Zumbrun. Symmetrizers and continuity of stable subspaces for parabolic-hyperbolic boundary value problems. *Discrete Contin. Dyn. Syst.*, 11(1):205–220, 2004.
- [15] G. Métivier, K. Zumbrun. Hyperbolic boundary value problems for symmetric systems with variable multiplicities. *J. Differential Equations*, 211(1):61–134, 2005.
- [16] D. Michelson. Stability theory of difference approximations for multidimensional initial-boundary value problems. *Math. Comp.*, 40(161):1–45, 1983.
- [17] R. Mneimné, F. Testard. *Introduction à la théorie des groupes de Lie classiques*. Hermann, 1986.
- [18] S. Osher. Systems of difference equations with general homogeneous boundary conditions. *Trans. Amer. Math. Soc.*, 137:177–201, 1969.
- [19] J. Ralston. Note on a paper of Kreiss. *Comm. Pure Appl. Math.*, 24(6):759–762, 1971.